The University of Queensland

Bachelor of Engineering Thesis

Dictionary Learning for Sparse Representation Based Image Fusion

Student Name: Tristan Burns 42648493
Course Code: MECH4500
Supervisor: Prof. David Mee
External Supervisors: PD Dr.-Ing.habil. Xiaoxiang Zhu
Dipl.-Math.techn. Claas Grohnfeldt
Submission Date: 31 October 2014

A thesis submitted in partial fulfilment of the requirements of the
Bachelor of Engineering Degree in Mechanical and Aerospace Engineering

UQ Engineering

Faculty of Engineering, Architecture and Information Technology

Dictionary Learning for Sparse Representation Based Image Fusion

by

Tristan Burns

A thesis submitted in partial satisfaction of the
requirements for the degree of
Bachelor of Engineering

in

Mechanical and Aerospace

with the

School of Mechanical and Mining Engineering

of

The University of Queensland

in cooperation with the

DLR German Aerospace Center

Committee in charge:

Prof. David Mee, Chair
Dipl.-Math.techn. Claas Grohnfeldt
PD Dr.-Ing.habil. Xiaoxang Zhu

Summer 2014

Dictionary Learning for Sparse Representation Based Image Fusion

Abstract

Dictionary Learning for Sparse Representation Based Image Fusion

by

Tristan Burns

Bachelor of Engineering in Mechanical and Aerospace

The University of Queensland, DLR German Aerospace Center

Prof. David Mee, Chair

Sparse representation based fusion of optical satellite images that have different spectral and spatial resolution is a rapidly growing research field. The central idea behind these methods is that patches of the fusion result, an image with high spectral and spatial resolution, have a sparse representation in a dictionary that is constructed from an input image with low spectral and high spatial resolution. Given the importance of this dictionary to the quality of the final high resolutions multispectral image, it is essential to use an intelligent dictionary selection or modification method. The pan-sharpening reconstruction performance for the state of the art J-SparseFI algorithm is investigated for the selection and training of coupled, local low and high resolution dictionaries, composed of corresponding low and high resolution panchromatic image patches. The performance is assessed for ten separate local dictionary selection methods, selecting coupled local dictionary atoms on the basis of distance, similarity or probabilistic dissimilarity to the current patch under reconstruction. Findings suggest an intriguing and counterintuitive tradeoff between spectral fidelity and spatial performance. Dictionary selection recommendations based spatial performance, robustness and spectral performance are made. A K-SVD based dictionary post training algorithm is also proposed. Modest performance improvements are observed when sharping a mutually uncorrelated WorldView-2 red-edge multispectral channel. Recommendations are made for future investigations into dictionary selection and training for J-SparseFI, with provision for extension into the Hyperspectral-Multispectral data fusion regime.

To Julie and Michael Burns

This thesis is dedicated to you both. My work and achievements are only possible through your love, wisdom, and constant support. For this I am unable to articulate the extent of my gratitude. With love, always, from your son,

Tristan

Contents

List of Figures

Acknowledgments

I would firstly like to acknowledge my close friends and family for their support and love throughout this endeavor.

I would like to thank the Deutsches Zentrum für Luft-und Raumfahrt (DLR - German Aerospace Center) for access to their resources, data and expertise. This work was supported by [Quote project money/grant/resources?? @Xiaoxiang This is normally included in thesis work]. I would especially like to thank PD Dr.-Ing.habil. Xiaoxiang Zhu and Prof. Richard Bamler and for providing me with a platform to write my thesis at the DLR German Aerospace Center.

I would also like to thank Prof. David Mee and Ms Rose-Marie Clements from The University of Queensland School of Mechanical and Mining Engineering, for their support in both the bureaucracy and realisation of this project.

Finally, a most sincere thanks goes to Dipl.-Math.techn Claas Grohnfeldt for his enduring guidance, patience, support and outstanding generosity and kindness. You are a true *Pfundskerl* and a great friend, and it has been a pleasure to work with you.

Introduction

Earth observation satellites typically generate multiple complementary products from respectively equipped sensors. In the case of optical earth observation satellites, such as IKONOS, Quickbird, GeoEye and WorldView-2, both a single broadband channel panchromatic (Pan) image and a multispectral (MS) image consisting of multiple spectral channels are acquired.

Pan images consist of high spatial resolution (HR) information with a typical ground sampling distance (GSD) of 0.3-1m, a significant technological achievement given that earth observation satellites typically orbit at an altitude of 600-800km. Pan images are thus suitable for accurate geometric analysis, however the single broadband channel provides inadequate spectral information for thematic interpretation. The insufficient spectral information limits possible insight into the material composition of the scene. Since the *single* panchromatic channel is responsive over a broad range of optical wavelengths, there exists no possibility for comparing multiple channel responses for various scene compositions, and limited intensity based estimation.

By contrast MS images consist of lower resolution (LR) spatial information, with GSDs on the scale of 2-5m. Such MS images, however, are applicable to thematic classification, as they are generally composed of 3-8 contiguous spectral bands over the visible-near infrared range (VNIR).

Thus given, data fusion techniques, aimed at extracting and combining the complementary spatial and spectral information from the respective pan and MS images, are under intensive development.

Image Fusion is the process of combining the relevant information from a set of images into a single image, such that the resultant fused image will be more informative and complete than the input images[1]. Of particular significance to this investigation is the process of combining the spatial information from a high resolution pancromatic image (PanHR) and the spectral information of a low resolution multispectral image (LRMS) to generate a representative high resolution multispectral (HRMS) product, known in the signals processing community as *pan-sharpening*.

Over the past two decades a considerable variety of pan-sharpening approaches have been implemented. Two of the dominant frameworks for pan-sharpening currently are *Component Substitution* and *Multi-Resolution Analysis* (MRA). Component substitution based techniques implement per-pixel transformations of the spectral channels, which is advantageous due to its simplicity. The LRMS images are spatially interpolated to the resolution of the PanHR image and transformed into a "colour" representation, with spatial details represented in a single channel. Pan-sharpening is then achieved by replacing this channel with the PanHR image, and applying the inverse transform. This general technique underlies methods such as the popular Intensity-Hue-Saturation (IHS), where the intensity channel is substituted [2], Brovey Transforms [3], Principal Component Replacement/Analysis (PCR/PCA) [4] and the Gram-Schmidt (GS) transform [5].

MRA based techniques implement channel-wise spatial transforms, rather than per-pixel channel transforms of component substitution [6]. The general methodology involves the generation of a HR image from the grayscale transformation of the panchromatic image, with matching global statistics to the LRMS channel. A spatial multiresolution transform is applied to the synthesized image and low frequency data in this decomposition is replaced by information derived from the LRMS image. The pan-sharpened image is subsequently obtained through the application of the inverse multi-resolution transform to the respective channels. Common implementations of MRA employ multi-scale wavelet transforms [7][8][9].

The growth in potential remote sensing applications continues to drive the development of increasingly sophisticated techniques. Recently, compressive sensing (CS) based approaches to pan-sharpening have displayed promising results. An initial successful attempt is described in [10] where multispectral image patches are assumed to have a sparse representation in a dictionary randomly sampled from comparable MSHR images.

Sparse representation based pan-sharpening was further exploited and improved in the Sparse Fusion of Images (SparseFI) algorithm [11], developed by the German Aerospace Centre (DLR). This method represents the MSLR measurement as a sparse linear combination of PanLR patches, arranged into a LR dictionary. In contrast to [10], SparseFI exploits the sparse representation of multispectral image patches in coupled dictionaries selected only from the panchromatic image at hand. This has the advantage that no HR multispectral images from other sensors are required. It has been demonstrated that the SparseFI algorithm also does not assume any spectral composition model of the panchromatic image and gives robust performance against even gross spectral model errors.

Currently under development at the DLR is a sophisticated state of the art algorithm, based on sparse reconstruction and compressive sensing theory, the "Jointly Sparse Fusion of Images" (J-SparseFI) algorithm [12][13]. J-SparseFI represents an advancement over the predecessor SparseFI algorithm, differing in its treatment of sparse reconstruction, by recovering the sparse reconstruction coefficients in adjacent MS channels simultaneously, and regularising the optimisation problem with a l_{1-2} mixed norm penalty, instead of the previously used l_1 norm, which encourages both sparsity and joint coefficients. This joint sparsity assumption is based on the observation that the reconstructed HRMS patches contain comparable spatial information in adjacent channels, albeit with different relative weightings. Hence HRMS reconstructions are likely to consist of the of geographically corresponding patches with different weightings. J-SparseFI algorithm exploits this notion and demonstrates leading

performance in the field of pan-sharpening [13]. Implementations of J-SparseFI reconstruct a HRMS image patchwise and subsequently tile the reconstructed patches to yield the complete HRMS image.

Within this framework of the high performance SparseFI and J-SparseFI algorithms, there exists no systematic study of the performance impacts of selecting or modifying the local coupled HR and LR Pan dictionaries. Currently, coupled dictionaries, local to the current patch undergoing reconstruction, are selected from the corresponding subset of nearest neighbor Pan patches. Given that the coupled HR and LR coupled dictionaries contain the information exploited in the reconstruction of the HRMS patches, there exists the potential for performance increase and additional insight into reconstruction quality by intelligently selecting these dictionary atoms. Specifically, the reconstruction quality for dictionaries selected on the basis of similarity of dissimilarity to the current patch undergoing reconstruction is of considerable interest, not just for possible performance benefits for J-SparseFI, but for the compressive sensing community at large. As such this investigation seeks to characterise the quality of the J-SparseFI HRMS reconstruction acquired using 10 distinct coupled dictionary selection methods.

Furthermore, the potential exists for more favorable dictionary representations, through appropriate algorithmic modification of the coupled dictionaries. Such dictionary training is commonly considered in the case that there is a training dataset available, that is representative of the signal requiring reconstruction. In this case, given that HRMS image is unavailable for dictionary training, the initial reconstruction provided by J-SparseFI can be used as a training signal for a post training operation. The well known K-SVD dictionary training algorithm [14][15], increases the representational ability of a dictionary by adapting the atoms on the basis of a singular value decomposition of an error signal computed relative to the training signal. K-SVD implementations have been used in similar post training operations, for the problem of pan-sharpening [16]. A post training operation for J-SparseFI is

proposed and tested, based on K-SVD adaptation, to determine whether such post training operations offer any reconstruction quality benefits.

2.1 Motivation

Optical remote sensors, such as IKONOS, Quickbird, GeoEye and WorldView-2, provide valuable information about the surface of the earth, with various civil applications, such as environmental monitoring, land-cover classification, disaster monitoring, water and agricultural management and weather forecasting. Given the broad applications for multispectral earth observation, any improvement in spectral fidelity or spatial accuracy of acquired earth observation satellite products is of considerable value to civilian, security and pure research interests.
The field of data fusion for satellite product enhancement, specifically pertaining to pan-sharpening of multispectral images and hyperspectral - multispectral image fusion, is one of the current research focuses at the Deutsches Zentrum für Luft und Raumfahrt (DLR).

Many such pan-sharpening and image fusion techniques exist. A recent approach to pan-sharpening under development at the DLR is the aforementioned J-SparseFI algorithm. For pan-sharpenning, this technique exploits a sparse representation of coupled HR and LR dictionaries composed of measured Pan patches. J-SparseFI is readily generalised to hyperspectral-multispectral image fusion, by replacing the PAN patch dictionaries with a MS patch dictionaries and sharpening bunches of hyperspectral (HS) channels [17].

Given the dependence on the coupled dictionary to appropriately represent the reconstruction, an investigation into the best methods to select or modify such a dictionary, in order to maximize reconstruction quality is desired. No systematic study of dictionary selection or training for J-SparseFI has been undertaken, hence an assessment and implementation of viable algorithms is valuable in the context of improving the performance of the J-SparseFI algorithm under development at the DLR.

Due to J-SparseFI being readily applicable to the problem of hyperspectral-multispectral

image fusion, dictionary selection and training results attained in examining pan-sharpening performance trials may yield insight and generalise to the hyperspectral-multispectral image fusion problem. As such, this work is being conducted as part of a larger body of work corresponding to the upcoming ($\sim$ 2017) launch of the German Environmental Mapping and Analysis Program (EnMAP) hyperspectral satellite. This mission aims to monitor and characterise the Earth's environment on a global scale by measuring key dynamical processes of the Earth's ecosystem by exploiting the high spectral resolution of hyperspectral imaging.

2.2 Scope and Goals

Before outlining the scope and specific goals of this project it is useful to define the following terms:

Remark 2.1. *In the context of this thesis a distinction is made between dictionary selection and dictionary training.*
Dictionary selection *is defined as the procedure by which a subset of predefined dictionary elements are chosen for inclusion in the reconstruction algorithm. With respect to the J-SparseFI algorithm, dictionary selection is the method by which HR and LR Pan patches are included in the HR and LR dictionaries, respectively.*

Dictionary training *is defined as the modification of the atoms of an existing dictionary, in an attempt to better represent the final reconstruction. With respect to the J-SparseFI algorithm, dictionary training involves modification of the HR and LR dictionaries, such that the individual atoms no longer necessarily represent the respective Pan patches.*

The primary outcomes of this work are characterisation, enhanced understanding and reconstruction quality improvements, with regards to various coupled dictionary selection methods and coupled dictionary training based on the K-SVD algorithm, for pan-sharpening using

the J-SparseFI algorithm.

The goals and deliverables of this work are broken down into categories, as follows:

- **Software Deliverables:** A variety of coupled dictionary selection methods are to be implemented in the high performance J-SparseFI c++ application. A dictionary training algorithm should also be implemented for the c++ application. All software must run on the Leibniz-Rechenzentrum (LRZ) SuperMUC supercomputing facilities, Garching, Munich.

- **Testing:** Coupled dictionary selection and training testing is to be conducted on realistic synthetic MSLR and PanHR data, synthesised from HRHS airborne HySpex data. Test reconstructions should be compared agains a synthetic MSHR reference image, to provide reconstruction quality assessment. All testing should be conducted using the SuperMUC supercomputing facilities.

- **Recommendations:** An in depth coupled dictionary selection and training analysis and critical evaluation is to be conducted. Recommendations based on reconstruction spatial accuracy, robustness with respect to differently sized dictionaries and spectral fidelity are to be made. Possible avenues of future work should be indicated.

2.3 Structure

This thesis is composed of chapters which are detailed as follows. To place this work in context, Chapter 3 provides some background on sparse representation based compressive sensing and introduces J-SparseFI and its predecessor SparseFI. Details regarding the synthesis of WorldView-2 data used for this investigation, reconstruction area assessed, as well as background for the supercomputing system employed for computation is provided as part of the testing overview in Chapter 4.

Following the introductory chapters, this thesis is arranged into two separate parts. Part 1 contains the dictionary selection half of this investigation. Ten dictionary selection methods are explained and results of testing are provided in Chapter 5. The results acquired in Chapter 5, for the ten selection methods, are critically analysed and evaluated in Chapter 6. Dictionary selection recommendations are also made in Chapter 6.

Part 2 covers the dictionary training component of this investigation. Chapter 7 provides an overview of the well documented K-SVD algorithm, and introduces the proposed post training algorithm for J-SparseFI. The results of K-SVD post training, for two separately selected dictionaries, are presented in Chapter 8. These results are evaluated and analysed in Chapter 9.

The significance of the overall findings and assessment of overall project goals is given in the conclusion and outlook chapter, Chapter 10, with a view to future work.

Image quality metrics used to characterise reconstruction performance in this investigation are outlined in Appendix A. Additional results for dictionary selection testing, including reconstructions and alpha values, are found in Appendix B. Further results obtained from the K-SVD post training investigation are documented in Appendix C.

Chapter 3

Sparse Representation Based Image Fusion

Recently there has been considerable interest in applying sparse representation techniques to the pan-sharpening problem. This chapter provides a concise literature review of sparse representation based compressive sensing (CS) and recent sparsity based pan-sharpening approaches. The high performance compressive sensing based algorithms under development at the DLR, namely ***Sparse Fusion of Images*** *(SparseFI) and* ***Jointly Sparse Fusion of Images*** *(J-SparseFI) are covered in detail, as they form the basis of this investigation. The prior art concerning problem parallelisation, and extension to the hyperspectral-multispectral (HSMS) image fusion for SparseFI and J-SparseFI is summarised. Crucially, the specific structure of the coupled HR and LR local dictionaries implemented in SparseFI and J-SparseFI will be introduced, such that the thesis topic of dictionary selection and training for sparse representation based image fusion is placed in specific context.*

3.1 Sparse Representation and Compressive Sensing

Many problems in image processing and remote sensing result in underdetermined inverse problems. If no additional knowledge is added to the system, it has, from a mathematical perspective, infinitely many solutions. In order to restrict the solution space to the physi-

cally meaningful solutions regularization is required. One increasingly popular regularization method is sparse reconstruction. In this case, the prior knowledge that the solution vector is sparse in some basis included in the system is exploited [15].

Compressive sensing takes advantage of the redundancy in many interesting signals, exploiting the fact that they are not pure noise. In particular, many signals are sparse, containing many coefficients close to or equal to zero, when represented in some domain.

A significant facet of the success of CS theory is that it can be performed with relatively efficient algorithms [18]. Since emphasis is placed on the undersampled case, the linear system describing the measurements is underdetermined, with infinitely many solutions. The sparse representation approach to compressive sensing employs regularization to constrain the system to a unique solution. Physically speaking, the main idea of sparsity assists in isolating the original or principal contributions to the signal [18].

In image processing applications, such as sparse reconstructions, a reliable image model is crucial for the recovery of representative products. Developments in the image processing community have revealed a tendency for natural images to be sparse within some basis spaces [19]. Given this, consider the sparse and redundant model of an image.

Consider a vectorised image, $\mathbf{x}, \in \mathbb{R}^n$, of size $\sqrt{n} \times \sqrt{n}$. Sparse representation theory assumes the existence of a matrix $\mathbf{\Phi} \in \mathbb{R}^{n \times m}$, with columns corresponding to a possible vectorised image. In the context of CS theory, $\mathbf{\Phi}$ is refered to as a **dictionary** and its columns are referred to as **atoms**. The image can thus be represented as:

$$\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha} \tag{3.1}$$

In the case that $\mathbf{\Phi}$ is overcomplete, i.e $\text{rank}(\mathbf{\Phi}) = n$ and $n < m$, there are infinitely many possible $\boldsymbol{\alpha}$ satisfying Equation (3.1). Given that a sparse $\boldsymbol{\alpha}$ is desired, with the fewest

nonzero elements, the problem can be formally represented as the following optimisation problem:

$$\hat{\boldsymbol{\alpha}} = \text{argmin}\|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{x}\|_2^2 = 0 \tag{3.2}$$

Where $\|\boldsymbol{\alpha}\|_0 = \#(i|\alpha_i \neq 0)$ corresponds to the number of nonzero elements in the vector $\boldsymbol{\alpha}$.

In the real world case of remote sensing, due to various physical and sensor limitations, $\mathbf{x}$ cannot be directly obtained. Instead we use an approximation of $\mathbf{x}$, which is given by the vectorised measurement image, $\mathbf{y} \in \mathbb{R}^n$, with, in this case, the same size as $\mathbf{x}$. The relation between the measured image and the original image can be expressed as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{3.3}$$

Where $\mathbf{A}$ is the sensing matrix, which can be interpreted as the encoding process in CS theory. CS theory ensures that under sparse regularisation [20] the image $\mathbf{x}$ can be approximately recovered as follows:

$$\hat{\boldsymbol{\alpha}} = \text{argmin}\|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 < \epsilon \tag{3.4}$$

The effectiveness of a sparsity prior as a regularisation term for ill posed problems is well documented [21][22][23][24]. However since $\|\boldsymbol{\alpha}\|_0$ is combinatorial, the optimization in Equation (3.4) is an NP-hard problem [19]. However under certain conditions for the matrix $\mathbf{D}$, it is shown that replacing the $l_0 - l_2$ minimization by $l_1 - l_2$ minimization yields the same solution with an overwhelming probability [25]:

$$\hat{\boldsymbol{\alpha}} = \text{argmin}\|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 < \epsilon \tag{3.5}$$

Where $\|\boldsymbol{\alpha}\|_1$ is the l_1 norm.

Definition 3.1. *The l_1 norm of a vector $\boldsymbol{\alpha}$ is defined as:*

$$\|\boldsymbol{\alpha}\|_1 := \sum_i |\alpha_i| \tag{3.6}$$

This corresponds to the sum of the absolute values of the elements in the vector $\boldsymbol{\alpha}$. It is also refered to as the Manhattan norm.

In order to deal with an unconstrained optimization problem, Equation (3.5) is commonly reformulated as the basis pursuit denoising (BPDN) optimisation problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \frac{1}{2} \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right\} \tag{3.7}$$

In which $\lambda||\boldsymbol{\alpha}||_1$ regularises the least square term $||\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}||_2^2$. Equation (3.7) is also referred to as the least absolute shrinkage and selection operator (LASSO) problem.

In practice, most specialised numerical algorithms can only solve the BPDN formulation. In modifying the parameter λ, exact congruence of $\mathbf{D}\boldsymbol{\alpha}$ and $\mathbf{y}$ is traded for sparsity in $\boldsymbol{\alpha}$. Successful application of sparse representation based compressive sensing relies on appropriately balancing the relative weight of reconstruction fidelity and sparse penalty regularisation.

This compressive sensing approach, featuring a sparse representation derived from BPDN optimisation is a feature of the SparseFI and J-SparseFI algorithms under development at the DLR. The dictionary representations of these algorithms form the basis of this investigations, thus the specifics of their implementation is introduced in Section 3.2 and Section 3.3.

3.2 SparseFI Algorithm

The underlying idea of the SparseFI algorithm is that the LRMS image can be represented, patch-wise, as a sparse linear combination of image patches extracted from the single channel low resolution panchromatic image (PanLR), which is obtained from appropriate filtering and downsampling of the observed PanHR image. This linear combination is then assumed to hold true for the HR system, which consists of an analogous dictionary of patches sampled from the PanHR image. The sparsity arises in accordance with the specific setup of the BPDN optimization problem and corresponding algorithm, consistent with the theory of sparse representation based compressive sensing [26].

The following SparseFI algorithm[1], as detailed in [26], attempts to reconstruct a HRMS image, $\mathbf{I}_Z$, by representing it as a sparse linear combination of patches of an observed PanHR image, $\mathbf{I}_{X_h}$, with $N_X = 1$ spectral channel[2]. The sparse coefficient, $\boldsymbol{\alpha}$, is obtained by optimizing an analogous LR system with information provided from the observed LRMS image, $\mathbf{I}_Y$, with $N_Y \in \mathbb{N}_+$ spectral channels:

SparseFI Algorithm

---

1. **Dictionary Learning**

   a) The PanHR image $\mathbf{I}_{X_h}$ is low-pass filtered and downsampled such that it has an identical sampling grid and similar final point spread function to the multispectral channels. The resulting PanLR image is denoted $\mathbf{I}_{X_l}$.

   b) The PanLR image, $\mathbf{I}_{X_l}$, and LRMS image, $\mathbf{I}_Y$, are tiled into $N_P \in \mathbb{N}_+$ partially overlapping patches, $\mathbf{X}_l^n \in \mathbb{R}^{P_l \times N_X}$ and $\mathbf{Y}^n \in \mathbb{R}^{P_l \times N_Y}$ respectively, where $n = 1, \ldots, N_P$ indexes the current patch and $P_l \in \mathbb{N}_+$ is the number of pixels in one vectorised LR patch.

   c) A local low resolution dictionary, corresponding to the *nth* patch undergoing reconstruction, $\mathbf{D}_l^n$, is constructed from a subset of the vectorised PanLR patches of $\mathbf{I}_{X_l}$, arranged column-wise. Likewise, a high resolution dictionary, $\mathbf{D}_h^n$, is constructed from the corresponding PanHR patches of $\mathbf{I}_{X_h}$. Note that before patches are added to the coupled high and low resolution dictionaries, $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$, the mean value of the patch is subtracted and then the patch is normalised by dividing each element by the l_2 norm of the HR patch. Hence, all dictionary atoms have zero mean and the HR dictionary atoms all have

[1]Note that the variables in [26] are denoted differently than those represented here.

[2] For the pan-sharpening N_X is always equal to 1. The notation is kept general for the extension to HS-MS image fusion, in which case $N_X > 1$ is the number of MS channels.

an l_2-norm equal to 1.

2. **Sparse Coefficient Estimation**

   a) This step attempts to optimise the LR system for a sparse representation of each LRMS patch, $\mathbf{y}_m^n \in \mathbb{R}^{P_l}$, in channel $m = 1, \ldots, N_Y$, as a linear combination of the PanLR patches, the atoms of the LR dictionary $\mathbf{D}_l^n$:

$$\mathbf{y}_m^n = \mathbf{D}_l^n \boldsymbol{\alpha}_m^n \tag{3.8}$$

   As the number of patches contained in the dictionary is much larger than the number of pixels in a LR patch, the dictionary is highly overcomplete, which means that the system is underdetermined. Therefore, it is argued that the "most probable" solution is the one employing the least number of Pan patches, a sparse representation. Note that prior to reconstruction, the mean value of the LRMS measurement vector, $\mathbf{y}_m^n$, is subtracted from the measurement vector and stored. It is later added to the HRMS reconstruction, $\mathbf{z}$, to ensure the reconstruction has the same mean value in each channel as the observed LRMS patch.

   b) The sparse representation coefficient, $\boldsymbol{\alpha}_m^n$ is computed via an $l_1 - l_2$ minimization, with the problem posed as follows:

$$\hat{\boldsymbol{\alpha}}_m^n = \arg\min_{\boldsymbol{\alpha}_m^n} \left\{ \lambda \|\boldsymbol{\alpha}_m^n\|_1 + \frac{1}{2} \|\mathbf{D}_l^n \boldsymbol{\alpha}_m^n - \mathbf{y}_m^n\|_2^2 \right\} \tag{3.9}$$

   Where λ is the regularisation coefficient of the sparsity inducing penalty term. Observe that the heavier the weighting of λ, the greater the extent the system tends to optimise for sparse solutions.

3. **HRMS Image Reconstruction**

a) The reconstructed HRMS image patches, $\mathbf{z}_m^n$, are *assumed* to share the same sparse coefficients as the corresponding LRMS image patches, $\mathbf{y}_m^n$.

b) Hence it is straightforward to reconstruct the final pan-sharpened HRMS image patch, by replacing the dictionary composed of PanLR patches, $\mathbf{D}_l^n$, with the corresponding dictionary composed of PanHR patches, $\mathbf{D}_h^n$:

$$\mathbf{z}_m^n = \mathbf{D}_h^n \hat{\boldsymbol{\alpha}}_m^n \tag{3.10}$$

Before the tiling, the mean value of the LRMS patch is added to $\mathbf{z}_m^n$, which, as a weighted sum of zero-mean patches, also has a mean value of zero. Subsequently tiling and summing all reconstructed patches, independently for each channel, yields the final pan-sharpened image, $\mathbf{I}_Z$.

3.3 J-SparseFI Algorithm

The fundamental idea behind the joint sparsity algorithm is the likely *correlation* between the relative values of the different multispectral channels [26]. The rationale behind this assumption is that the representation of the geometrical shapes of objects in the image scene will likely be present in each of the multispectral bands, albeit with different relative intensities.

Within the sparse representation regime, this assumption postulates that the different channels of the reconstructed images share the majority of their non zero components i.e. shared or joint sparsity. Crucially, testing has demonstrated that J-SparseFI outperforms SparseFI, in addition to a number of competitor algorithms, providing validation for the assumption of joint sparsity.

The J-SparseFI algorithm represents a modification of the SparseFI algorithm, and shares the Dictionary Learning step. The primary differences lie in the sparse coefficient estimation

step. This algorithm is also fully detailed in [26].

The following norm definitions are useful:

Definition 3.2. *The Frobenius matrix norm, denoted by* $\|\cdot\|_F$*, is defined for a matrix* $\mathbf{A} \in \mathbb{R}^{n\times m}$ *as:*

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} |\alpha_{ij}|^2} \tag{3.11}$$

Definition 3.3. *The* $l_{2,1}$ *combined norm of a matrix* $\mathbf{A}$ *is defined as:*

$$\|\mathbf{A}\|_{2,1} = \sum_{i=1}^{N} \|\boldsymbol{\alpha}_i\|_2 = \|(\|\boldsymbol{\alpha}_1\|_2, \ldots, \|\boldsymbol{\alpha}_N\|_2)\|_1 = \sum_{i=1}^{N} \left(\sum_{j=1}^{M} \alpha_{i,j}^2\right)^{1/2} \tag{3.12}$$

The mixed norm, $\|\cdot\|_{2,1}$*, represents the sum of the* l_2 *norms of the rows of a matrix.*

J-SparseFI Algorithm

---

1. **J-SparseFI Sparse Coefficient Estimation**

   a) The joint sparse representation is constructed by arranging the measurements and sparse coefficients, in individual channels, column by column, to form the following matrices:

$$\mathbf{Y}^n = \left[\mathbf{y}_1^n \ldots \mathbf{y}_{N_Y}^n\right] \tag{3.13}$$

$$\mathbf{A}^n = \left[\boldsymbol{\alpha}_1^n \ldots \boldsymbol{\alpha}_{N_Y}^n\right] \tag{3.14}$$

$$\mathbf{Z}^n = \left[\mathbf{z}_1^n \ldots \mathbf{z}_{N_Y}^n\right] \tag{3.15}$$

   Yielding the following low resolution system representation:

$$\mathbf{Y}^n = \mathbf{D}_l^n \mathbf{A}^n \tag{3.16}$$

b) Simultaneous recovery of *all* coefficients is facilitated by $l_{2,1}-l_2$ minimization, as follows:

$$\hat{\mathbf{A}}^n = \arg\min_{\mathbf{A}^n} \left\{ \lambda' \|\mathbf{A}^n\|_{2,1} + \frac{1}{2}\|\mathbf{D}_l^n \mathbf{A}^n - \mathbf{Y}^n\|_F \right\} \tag{3.17}$$

As such, the $l_{2,1}$ norm promotes sparsity along the columns of the matrix, while simultaneously minimizing the energy along rows. This corresponds to non-zero coefficients in the multispectral channels *that occupy the same rows* being favoured.

c) The final pan-sharpened HRMS patches are obtained by substituting in the HR dictionary, as with the SparseFI algorithm:

$$\mathbf{Z}^n = \mathbf{D}_h^n \hat{\mathbf{A}}^n \tag{3.18}$$

Which are subsequently summed to obtain the HRMS image, $\mathbf{I}_Z$.

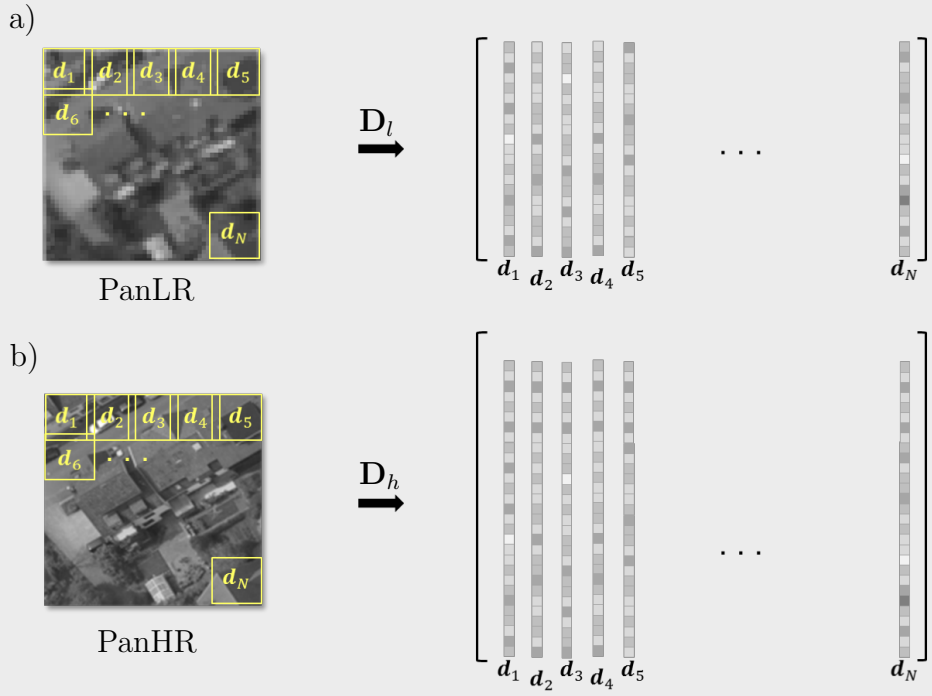
The J-SparseFI algorithm is readily generalised for Hyperspectral Multispectral (HM) image fusion, by utilizing a HRMS image in place of the PanHR image, and subsequently accounting for multiple channels in the sharpening image, $N_X \in \mathbb{N}_+$. Bunches of hyperspectral channels are then sharpened with multispectral channels with a relatively close spectral profile. This approach is discussed at length in [17].

Given the posing of the joint sparse reconstruction problem in J-SparseFI, the nature of the dictionary selection and training under investigation in this work is clarified. The coupled dictionary structure, implemented in both SparseFI and J-SparseFI algorithms, leverages the geometric information present in the PanHR image, by representing the HRMS reconstruction as a sparse combination of these elements. The weighting of this combination is determined by solving the coupled LR system, consisting of the PanLR image and MSLR measurement. The dictionary in the LR system consists of patches corresponding to the HR dictionary, hence the dictionaries are referred to as coupled. This coupled dictionary

structure is depicted in Figure 3.1.

In order to limit the size of the sparse optimisation problem, it is vital to limit the size of the coupled dictionaries. It is thus sensible to select local coupled dictionaries, specific to the current patch under reconstruction, via some user defined method. There is also the possibility of algorithmically modifying the dictionaries, such that they better represent the signal under reconstruction. Respectively, such dictionary selection and dictionary training approaches form the basis of this investigation.

Figure 3.1: Here the fundamental coupled dictionary structure for SparseFI and J-SparseFI is depicted. The complete set of patches, displayed here, forms global dictionaries. The selection and training of local dictionaries, consisting of a subset of the complete set of patches, is used in this investigation, to limit the size of the optimization problem and possibly improve the quality of the fusion result.



Chapter 4

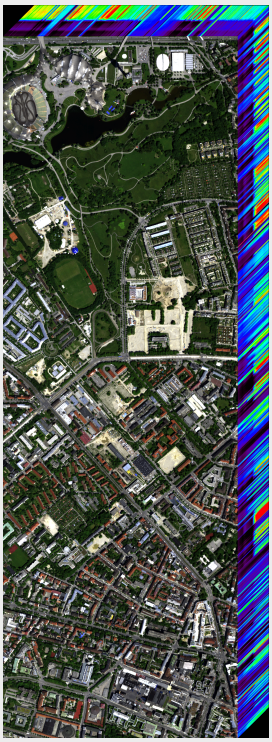
Testing Overview

An overview of the test setup is provided in this chapter. WorldView-2 images are synthesised from a hyperspectral data cube, collected from an airborne HySpex sensor. Details are provided regarding the Leibniz-Rechenzentrum (LRZ) SuperMUC supercomputing facilities used to compute representative reconstructions in this investigation. A representative sub-region of the complete image is detailed, for use in performance and quality testing.

4.1 Data Sets

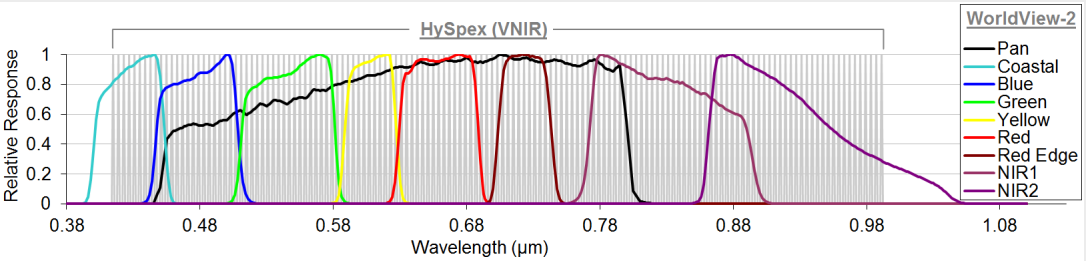
To provide justified evaluation regarding the performance of reconstruction algorithms, it is common to work with simulated data [13]. All dictionary selection tests are conducted using synthetic WorldView-2 images, simulated from airborne VNIR HySpex data aquired over Munich, Germany, in 2012. The HySpex sensor achieves $\approx 1m$ ground sampling distance (GSD), when flown at an altitude of 1000m, across 160 spectral channels spanning wavelengths from 0.4 to $1.0\mu m$.

Figure 4.1: Here the hyperspectral datacube collected from an airborne HySpex sensor, over Munich, Germany. The 160 spectral channels are selectively filtered to generate synthetic WorldView-2 testing and reference MS and Pan images.



The collected hyperspectral data cube is displayed in Figure 4.1 and has a spatial resolution of 1m and a size of 3600×1200 pixels. The relative spectral responses of both HySpex and WorldView-2 are shown in Figure 4.2.

Figure 4.2: Here the grey peaks denote the spectral response of the VNIR HySpex Sensor, while the coloured lines denote the relevant WorldView-2 spectral responses for the MS channels. The black line denotes the spectral response for the WorldView-2 Pan channel.



Using the hyperspectral data cube as input, synthetic MSLR, PanHR and PanLR images were developed conforming to the spectral and spatial specifications of the WorldView-2 imager. Additionally a reference MSHR image was generated, to provide a benchmark for reconstructed MSHR image comparison. The reference MSHR image and PanHR image are of the same spatial resolution as the hyperspectral data, 1m.

PanLR and MSLR images were synthesised by low pass filtering and downsampling the respective MSHR and PanHR images. While typical topographic satellites operate with a spatial resolution ratio between the PanHR and MSLR image of $F_{DS} = 4$, this investigation uses a more extreme resolution ratio of $F_{DS} = 10$, to investigate the impacts of dictionary selection and training with data sets approaching the limits of the pan-sharpening methods.

4.2 Supercomputer Architecture and Data

Considering the number of tests desired for both dictionary selection and training, the dataset size and requisite number of reconstruction patches, and the application of numerically ex-

pensive methods such as convex optimisation and singular value decomposition, high computational power was required for the acquisition of representative results on realistic datasets. As such, all testing was performed on the Leibniz-Rechenzentrum (LRZ) SuperMUC supercomputing facilities in Garching, Munich.

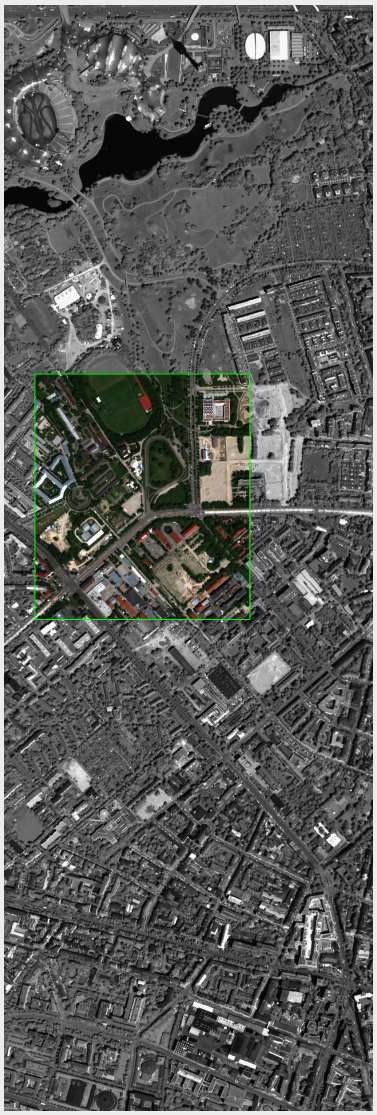
With a peak performance of 3.2 PFLOPS, the x86-based SuperMUC is one of the world's fastest supercomputers, ranking tenth in the Top500 List as of November 2013 [27]. It is built out of 18432 Intel Xeon E5-2680 processors running at 2.7GHz, which sums up to a total of 147,456 cores [28]. One compute node of SuperMUC consists of 2 sockets, each equipped with 8 cores. A single node can access 32 GiB of main memory, resulting in 2GiB per core and 288TiB for the entire machine.

In order to process large scale Earth observation data, the J-SparseFI image fusion software is optimised for memory exploitation on the SuperMUC. Internode MPI communication is kept low in order to maximize parallel efficiency. The J-SparseFI application is compiled with the Intel® icip compiler using the system's wrapper mpiCC.

4.3 Reconstruction Area

In order to preserve valuable cpu project hours, a sub-region of the complete image is reconstructed to MSHR. This reduces computational time proportionally, by limiting the number of regularised sparse optimization problems that require solving. However, the *key* to ensuring that the dictionary selection and training is representative is to select patches from the *complete* PanHR and PanLR images, for the respective coupled HR and LR dictionaries, $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$. This ensures that the dictionaries obtained would be identical to the dictionaries selected for reconstruction of the entire image.

Figure 4.3: This image displays the area that is reconstructed to a HRMS image for dictionary selection and training, outlined in green. Note that to provide representative coupled dictionary testing, both the HR and LR dictionary atoms are selected from the complete respective PanHR and PanLR images, as indicated by the surrounding PanHR image in the figure.



The sub-region is an 800×700 pixel region, indicated in Figure 4.3, that contains a representative combination of artificial structures, such as roads and buildings, and vegetation, such as grasses and trees. This sub-region is detailed in Figure 4.4.

Figure 4.4: Here close up of the reconstructed sub-region is provided for a) the reference HRMS image and b) the input LRMS image, to be reconstructed to HR. Note the combined presence of urban structures and vegetation.

4.4 J-SparseFI Settings and System Parameters

In all tests conducted in this work, the regularization parameter λ is set to 1. A separate sensitivity study on lambda for J-SparseFI using the NNP dictionary selection method has revealed that the sensitivity of the reconstruction quality with respect to lambda is very little for lambda values between 10^{-1} and 10^{1} [13].

The size of the squared LR patches is set to $P_l = 5 \times 5 = 25$. Given that the HR-LR resolution ratio is equal to 10/1 pixels in both spatial directions, the HR patches consists of $P_h = 50 \times 50 = 2500$ pixels.

Remark 4.1. *A technical note regarding patch location. Patches are linearly numbered from the top-left corner to the bottom-right corner. The set of vectorized LR/HR patches are denoted as* $\mathbf{x}_l^i/\mathbf{x}_h^i$, $i = 1, ..., N_P$. *Within the image, patches have two coordinates, one horizontal,* $u = 1, ..., N_{P_U}$, *and one vertical,* $v = 1, ..., N_{P_V}$, *with* N_{P_U} *indexing the total number of patches in the vertical direction and* N_{P_V} *indexing the total number of patches in the horizontal direction.* N_{P_U} *and* N_{P_V} *depend on the image size, the downsampling factor* F_{DS}, *the patch size and the overlap. The total set of patches is then* $N_P = N_{P_U} N_{P_V}$.

The linear index for each patch is computed as $i = (u-1)N_{P_V} + v$. *The 2D patch coordinates are denoted as* $\mathbf{q}_i = [u, v]$, *where* $u = \lfloor (i-1)/N_{P_V} \rfloor + 1$ *and* $v = i - (u-1)N_{P_V}$.

The downsampling factor used in testing was $F_{DS} = 10$. *For this investigation the patch size was chosen to be* $5 \times 5 = 25$ *LR pixels,* $50 \times 50 = 2500$ *HR pixels and that the patch overlap is 1 LR pixel = 10 HR pixels in both horizontal and vertical direction. As* $F_{DS} = 10$, *we have* $N_{P_U} = 356$ *and* $N_{P_V} = 116$, *resulting in* $N_P = 41296$ *patches available for selection in all tests.*

Part I

Dictionary Selection

Chapter 5

Dictionary Selection Implementations For J-SparseFI

Essential to the J-SparseFI pan-sharpening algorithm is the coupled HR-LR Pan dictionary structure, outlined in Figure 3.1. Appropriate selection of PanHR and PanLR patches for inclusion in the coupled dictionary structure presents the possibility for performance enhancement, in terms of spectral and spatial metrics. This chapter details 10 non-local dictionary selection methods, and compares reconstruction performance with coupled dictionaries built up from a set of selected nearest neighbor patches, which is currently the standard technique detailed in the literature concerning J-SparseFI pan-sharpening and hyperspectral-multispectral image fusion. Testing is conducted using simulated WorldView-2 data, with calculations performed on the Leibniz-Rechenzentrum (LRZ) SuperMUC supercomputing facilities in Garching, Munich. Crucially, an apparent trade off between spectral fidelity and spatial accuracy is revealed in the selection methods assessed. This presents intriguing possibilities for extension to coupled dictionary selection for the HS-MS fusion domain.

5.1 Coupled Dictionary Selection

Consistent with implementations in compressive sensing, such as the aforementioned SparseFI and J-SparseFI algorithms, is the representation of the model based reconstruction as a lin-

ear combination of a set of elements, in this case patches from the HR and LR Pan images. Within the signal analysis community the elements are referred to as *atoms* and the set is referred to as a *dictionary*.

In the context of this thesis dictionary selection is defined as follows.

Remark 5.1. ***Dictionary selection*** *is defined as the procedure by which a subset of predefined dictionary elements are chosen for inclusion in the reconstruction algorithm. With respect to the SparseFI and J-SparseFI algorithms, dictionary selection is the method by which HR and LR Pan patches are respectively included in the HR and LR dictionaries.*

Given the considerable computational demands resulting from the large number of regularised sparse coefficient optimisation problems inherent in the J-SparseFI algorithm, it is essential to select coupled dictionaries of appropriate size. While it is theoretically feasible to construct coupled dictionaries consisting of the *complete* set of patches from the HR and LR Pan images, this would result in excessive computational requirements and, as revealed by experimentation, lower quality reconstructions. Hence, with respect to image quality there is an optimal dictionary size depending on both the dictionary selection method and the underlying data. The complexity of the individual optimization problem on the other hand, monotonically increases with the number of dictionary atoms.

As such, local coupled dictionaries containing N_{DP} atoms of corresponding PanHR and PanLR image patches are constructed:

$$\mathbf{D}_l^n \in \mathbb{R}^{P_l \times N_{DP}} \tag{5.1}$$

$$\mathbf{D}_h^n \in \mathbb{R}^{P_h \times N_{DP}} \tag{5.2}$$

Where P_h is the size of the vectorised PanHR image patch and P_l is the size of the vectorised PanLR image patch. Note that the dictionaries implemented in both SparseFI and

J-SparseFI are local to the *current* patch undergoing reconstruction, and hence the coupled dictionaries change dynamically according to the patch in question, indicated by the dependency on $n = 1, \ldots, N_P$.

It is important to note, prior to selection and inclusion in the HR or LR dictionaries, all dictionary atoms have their means subtracted and are then normalised through division with their l_2 norm. Consider Equation (5.3) and Equation (5.4), the mean subtraction and normalization for the nth vectorised PanHR image patch, $\mathbf{x}_h^n$, prior to dictionary selection:

Mean Subtraction

$$\mathbf{x}_h^n - \bar{\mathbf{x}}_h^n \tag{5.3}$$

Normalisation

$$\frac{\mathbf{x}_h^n}{\|\mathbf{x}_h^n\|_2} \tag{5.4}$$

Since all PanHR patches in the dictionary are normalized, the coefficients of the reconstructed coefficient vector $\boldsymbol{\alpha}^n$ do not depend on the relative illumination of one patch. Therefore the coefficients in $\boldsymbol{\alpha}^n$ reflect the actual contribution of each dictionary atom to the reconstruction result $\mathbf{Z}^n$, significantly facilitating interpretation. This procedure presents the additional benefit of emphasising geometrical features present in the images, permitting greater discrimination in similarity based dictionary selection techniques.

5.2 Nearest Neighbor Patches (NNP) Dictionary Selection

5.2.1 Overview

Currently, both the pan-sharpening [12] [26] and HS-MS fusion [17] implementations of J-SparseFI employ local coupled dictionaries selected from a subset of N_{DP} nearest neighbor

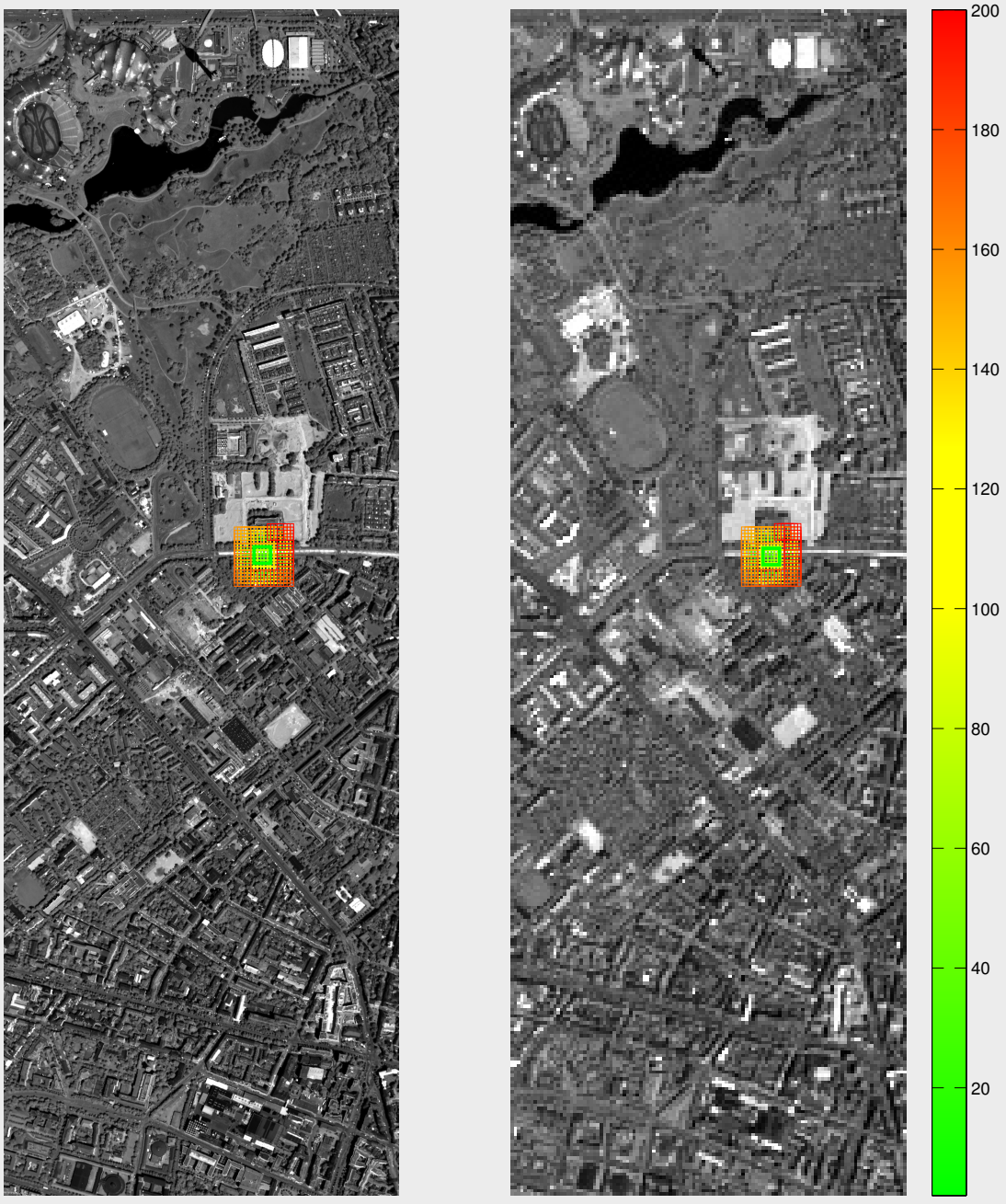
patches to the current patch under reconstruction. The motivation behind this metric is two-fold: Firstly, it is assumed that patches in the neighborhood contain similar or comparable spectral and spatial information, and secondly, the selection metric has relatively low computational demands. Though robust J-SparseFI performance has been demonstrated using this metric, it by no means precludes possible improvements resulting from more sophisticated methods. Given its current usage, the Nearest Neighbor Patches (NNP) dictionary selection method will be referred to as the **benchmark**, for comparison against the other selection methods under investigation.

In this method, all PanLR and corresponding PanHR patches are assigned coordinates $\mathbf{q}$, as introduced in Remark 4.1. To build up the local dictionary for the current patch i under reconstruction, the maximum norm of the distance between patch i and the other $n = 1, \ldots, N_{DP}$ patches are evaluated respective to their coordinates:

$$\mu_{NNP}^{i,n} := \|\mathbf{q}_i - \mathbf{q}_n\|_\infty \tag{5.5}$$

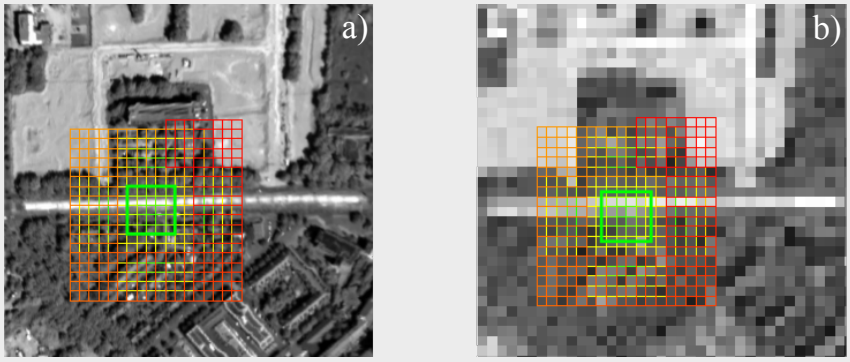
The N_{DP} closest patches according to this maximum norm metric, $\mu_{NNP}^{i,n}$, are included in the respective HR and LR coupled dictionaries. This results in patches being selected from a characteristic square surrounding the current patch undergoing reconstruction.

Figure 5.1: The NNP coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the maximum norm, i.e. μ_{NNP}.



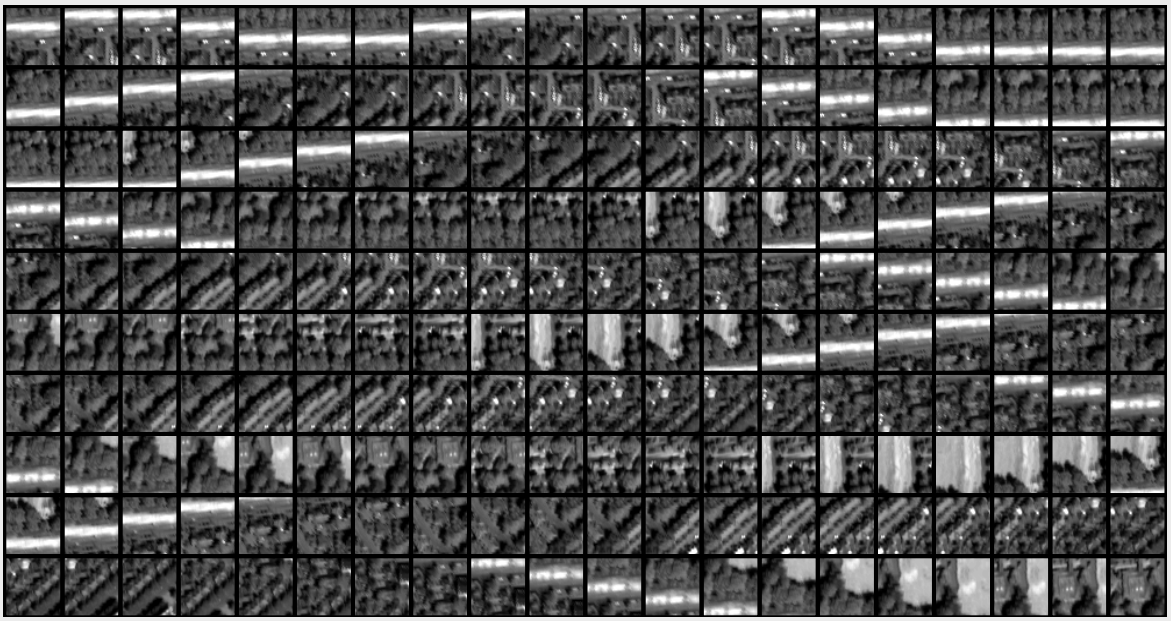
The example road patch in Figure 5.1 is chosen for qualitative comparison between the non-local dictionary selection methods. It is suitable for comparison since it provides well defined geometric information, in the form of the road, and a combination of vegetation and synthetic material. It thus is able to provide insight into the workings of the patch selection algorithms employed in the non-local methods under investigation. A detailed view of the area surrounding the example road patch is shown in Figure 5.2.

Figure 5.2: Detailed view of the local surroundings of the example road patch in a) HR and b) LR, which will be used for qualitatively comparing dictionary selection methods under investigation.



A detailed visualisation of the high resolution dictionary, $\mathbf{D}_h^n$, for the NNP selection method is presented in Figure 5.3. The dictionary is ordered according to the respective maximum distances from the patch under reconstruction.

Figure 5.3: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NNP selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



In Figure 5.3 the road geometry of the example patch under reconstruction is more visible. Note also the presence of vegetation on either side of the road. The contrast, geometry and scene variety of this patch makes it valuable as a qualitative comparison point.

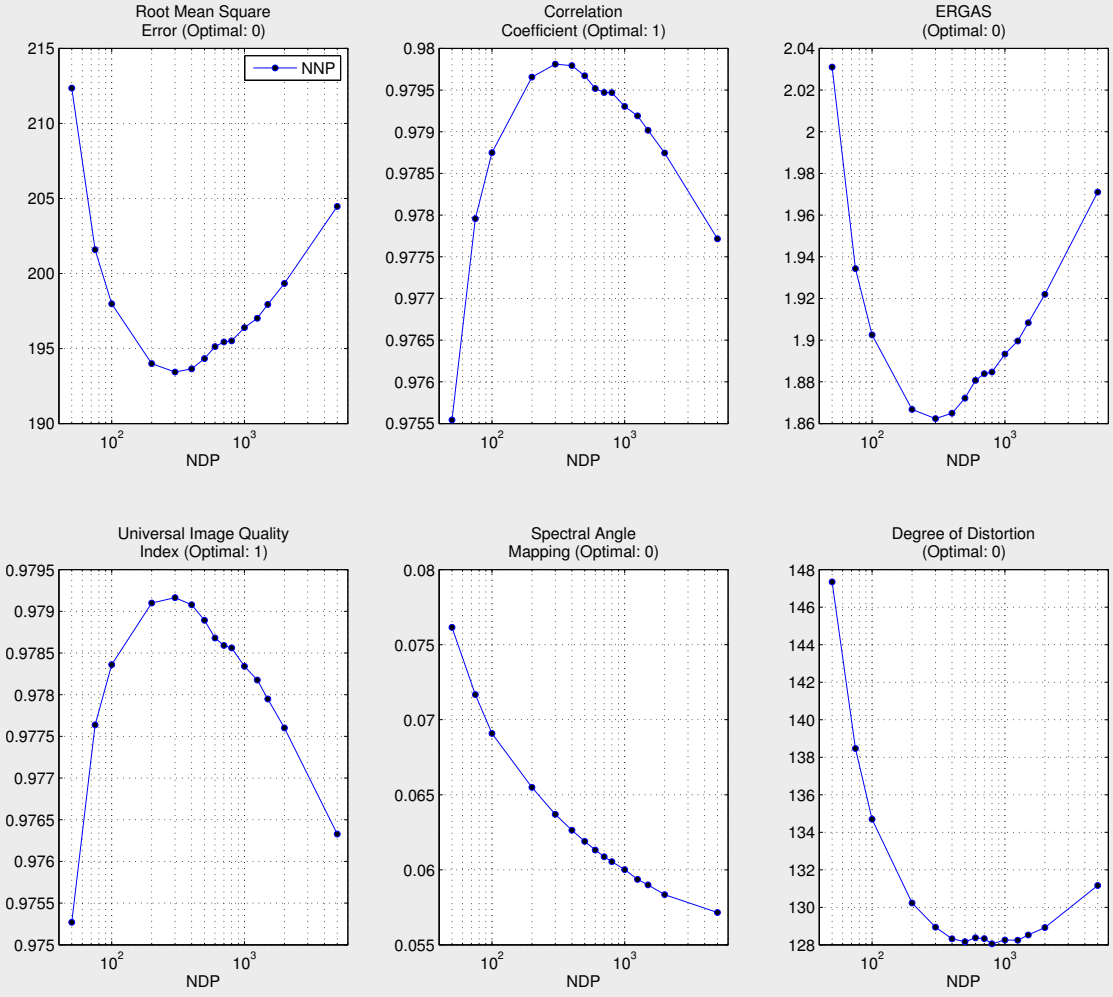
Periodicity is apparent in the example dictionary patches, a direct result of the maximum norm selecting rings of patches at progressively increasing distances from the patch under reconstruction. Also evident is visual scene similarity, the underlying assumption of this selection method, though as a consequence of the inherent periodicity, positioning of geometrical features, such as the road, is predominantly offset with regards to the patch under reconstruction.

5.2.2 NNP Benchmark Performance

The performance of the J-SparseFI algorithm on the given reconstruction area, depicted in Figure 4.3, is assessed in terms of standard image quality metrics, detailed in Appendix A, for the benchmark NNP dictionary selection method. The results of the quantitative testing is presented in Figure 5.4, for a range of dictionary sizes, from $N_{DP} = 50$ to $N_{DP} = 5000$.

All metrics are computed by comparison with the synthetic reference HRMS image (see Section 4.1), as per Appendix A.

Figure 5.4: The performance of the benchmark NNP dictionary selection method is assessed for varying N_{DP}. Image quality metrics are detailed in Appendix A



With the exception of the Spectral Angle Mapping (SAM), which purely measures spectral information preservation, all metrics are optimised for an $N_{DP} \approx 300$. This provides an initial indication that a trade off may exist between spectral fidelity and geometrical accuracy.

5.3 Non-Local PanLR Norm Comparison Dictionary Selection (NL1)

5.3.1 Overview

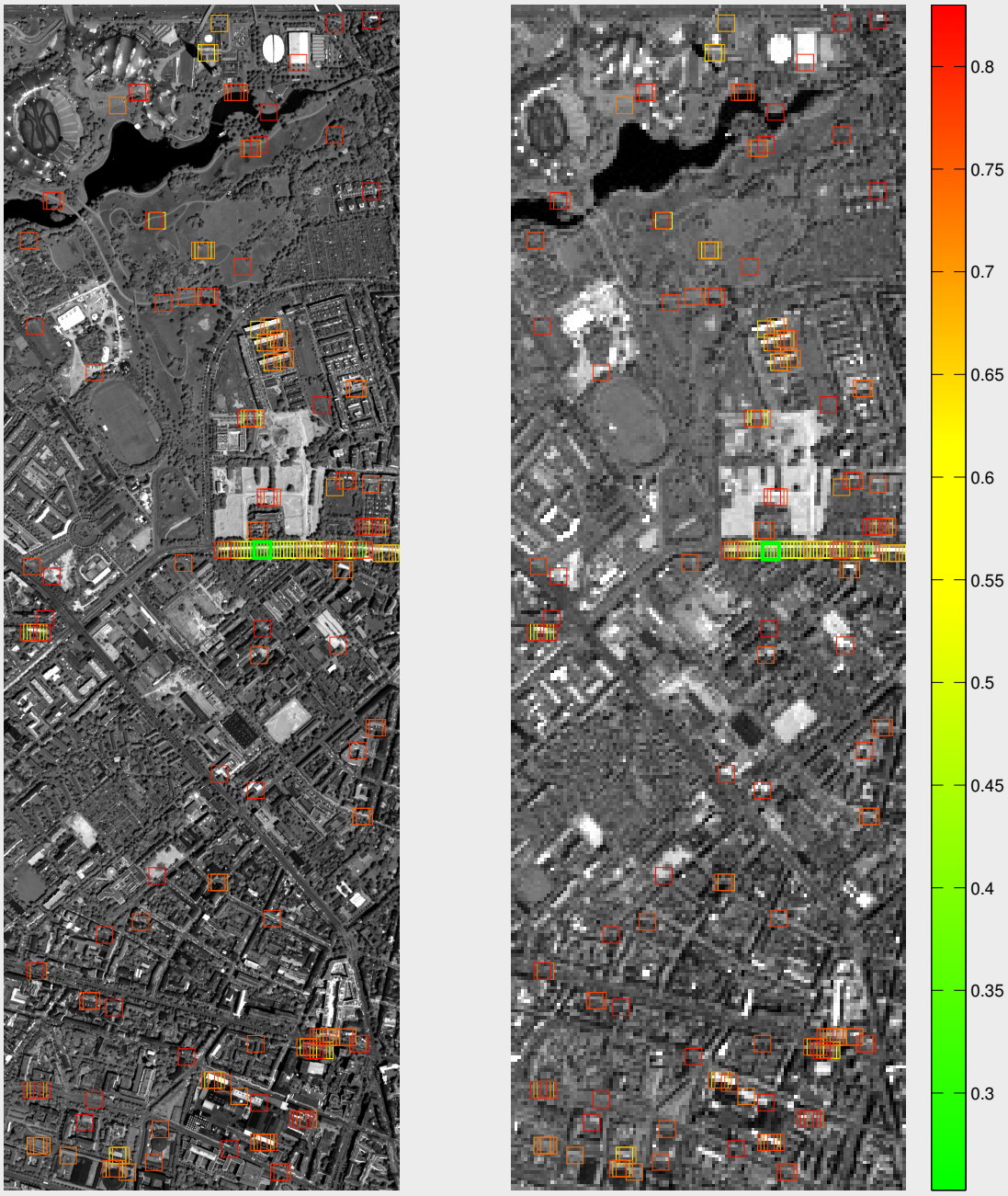
The first Non-Local dictionary selection method under investigation is based upon selecting patches which are most similar to the *ith* patch under reconstruction, in terms of PanLR norm comparison. Essentially, the euclidean norm of the difference between the *ith* vectorised PanLR patch under reconstruction, $\mathbf{x}_l^i$, and the *nth* patch, $\mathbf{x}_l^n$, from the complete set of N_P patches, is calculated:

$$\mu_{NL1}^{i,n} := \|\mathbf{x}_l^i - \mathbf{x}_l^n\|_2 \tag{5.6}$$

The N_{DP} most similar patches according to this metric, $\mu_{NL1}^{i,n}$, correspond to the N_{DP} patches with the lowest euclidean norm difference in the LR system. These patch coordinates are subsequently taken and used to build the HR and LR coupled dictionaries.

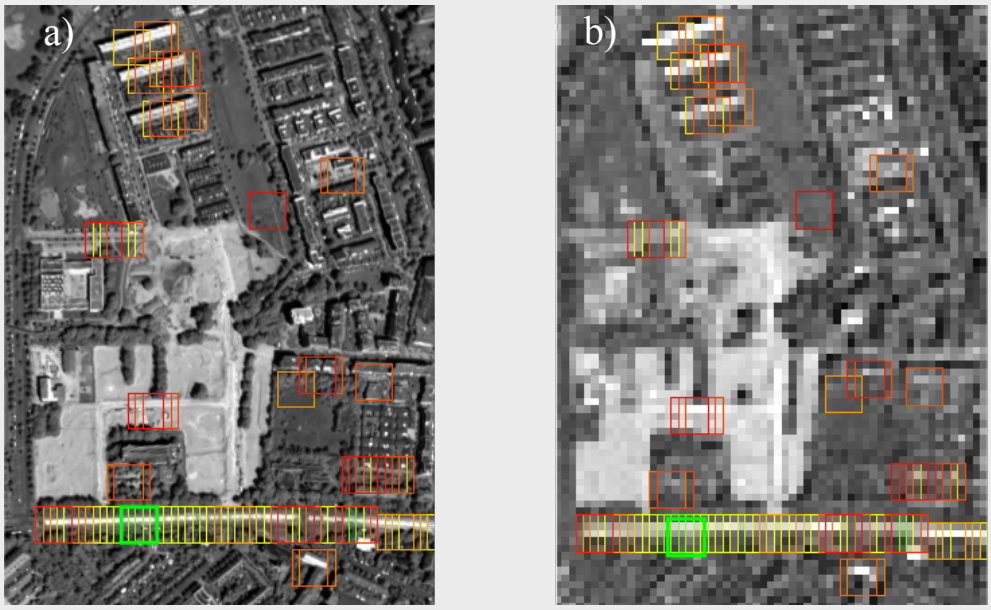
The NL1 dictionary location image is displayed in Figure 5.5.

Figure 5.5: The NL1 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the euclidean norm of the PanLR patch difference, i.e. μ_{NL1}.



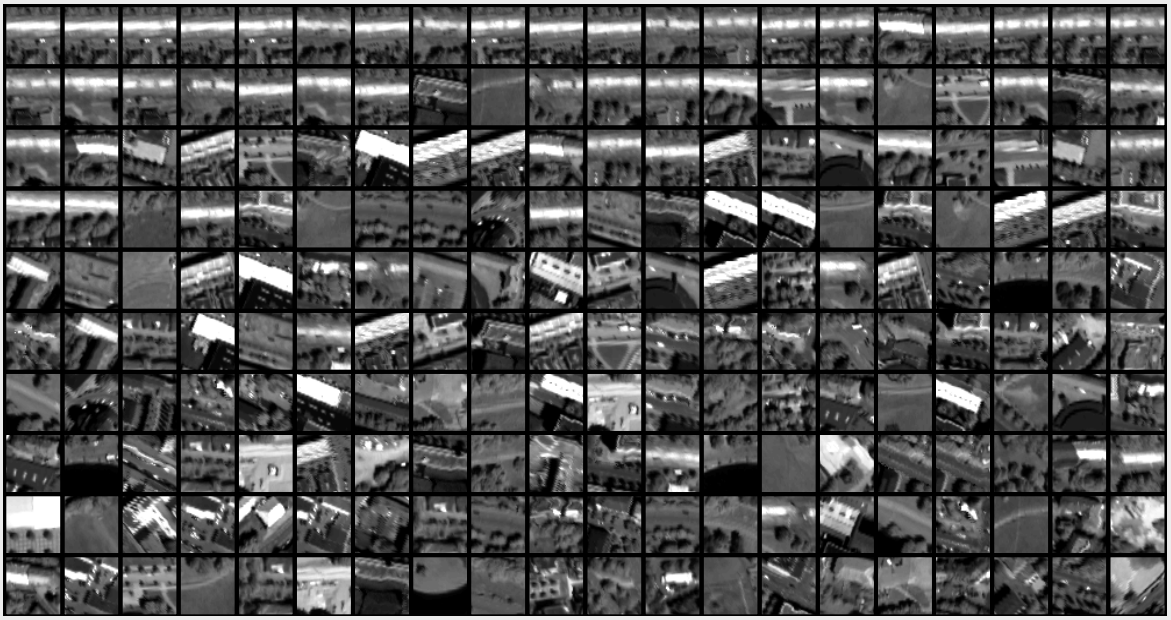
Notable in Figure 5.5 is the selection of all road patches, as would be expected from this selection metric, since the metric selects similar patches in the LR system. The selection of road patches can be better observed in the detail figure Figure 5.6.

Figure 5.6: Here the NL1 selection of similar road patches is displayed in detail for a) PanHR and b) PanLR images.



The prevalence of yellow indicators in Figure 5.5 signifies that many of the N_{DP} selected patches in the dictionary are not ranked at the extremity, in terms of LR euclidean norm distance. This is consistent with the loss of detail comparison involved with comparing LR patches. Also note the selection of the angled roof patches in Figure 5.6, which are selected on the basis of their angled white roof, which roughly approximates the road geometry.

Figure 5.7: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL1 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.

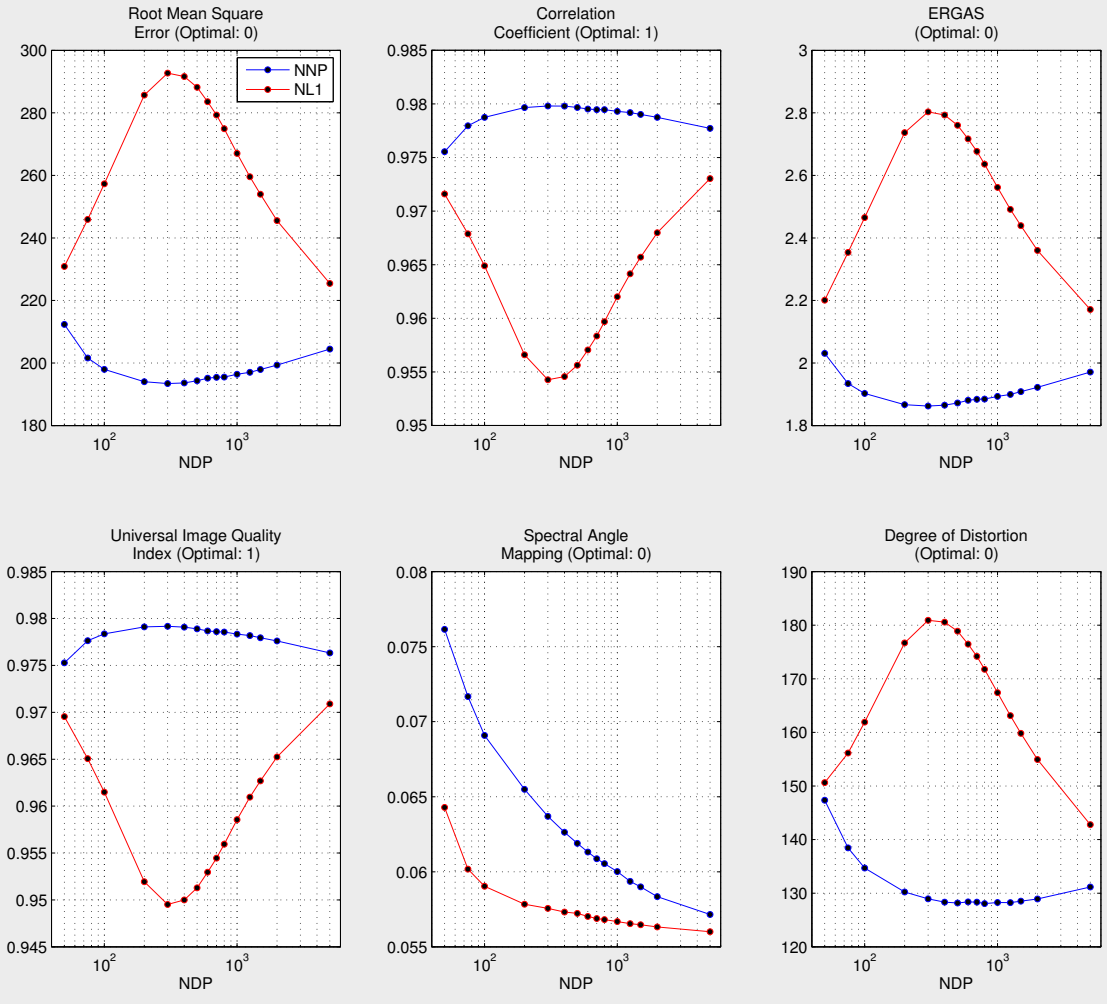


The detail view of the NL1 HR dictionary, $\mathbf{D}_h^n$, in Figure 5.7 further elucidates the dictionary structure for the example road patch. Patches with corresponding geometry, overwhelmingly other sections of the road, are favoured. Additionally, patches of high contrast, corresponding to lines approximating the road location, such as patch row 8, column 2, Figure 5.7. It is important to note that for the visualization of the dictionary in Figure 5.7, the patches are shown with their relative illumination preserved, i.e. before the mean subtraction and normalization. However, NL1 and all other dictionary selection methods work with the normalized zero-mean patches. Therefore all HR dictionary patches visualized in Figure 5.7 actually have an l_2 norm equal to one and a mean value equal to zero. Therefore, it is no surprise to find both very dark and very bright patches in the figure.

5.3.2 NL1 Performance

Quantitative performance testing was conducted on the given reconstruction area, displayed in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.8.

Figure 5.8: The performance of the NL1 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



It is evident in Figure 5.8 that the NL1 dictionary selection method is dramatically inferior to the benchmark NNP for all image quality metrics, with the exception of the spectral angle mapping, for which NL1 offers significant improvement for all N_{DP}. This is further evidence of the tradeoff between spectral fidelity and geometric correlation. Intriguingly the worst NL1 performance occurs for $N_{DP} \approx 200$, corresponding to the optimal NNP performance. An explanation for the reduction in quality is the loss of detail from HR to LR, weakening the assumption that the sparse coefficient is applicable to both HR and LR systems. This is discussed in greater detail in the Critical analysis and Evaluation, in Chapter 6.

5.4 Non-Local MSLR Norm Comparison Dictionary Selection (NL2)

5.4.1 Overview

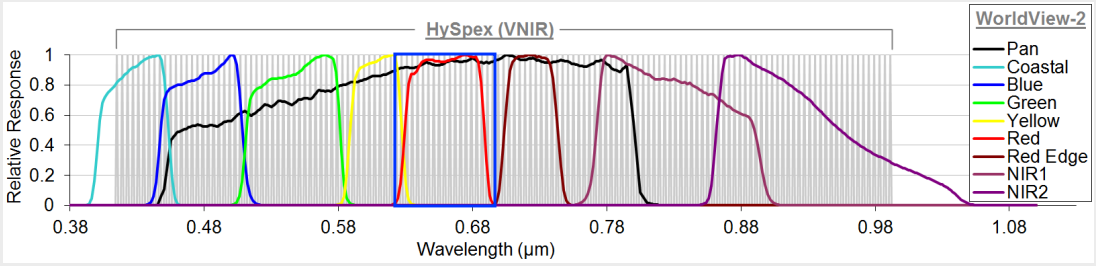
The underlying idea for the Non-Local MSLR Norm Comparison Dictionary selection method (NL2) is the possibility of improving reconstruction fidelity through the inclusion of information contained in the MSLR measurement image, $\mathbf{y}_m^n$. For this method it is assumed that a weighted linear combination of the $m = 1, \ldots, N_Y$ MSLR measurement channels can be used to form an approximate PanLR patch, $\mathbf{x}_{l,approx}^n$:

$$\mathbf{x}_{l,approx}^n := \sum_{m}^{N_Y} w_m \mathbf{y}_m^n \tag{5.7}$$

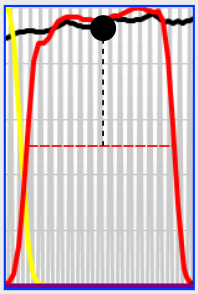
A similar assumption is made in [10]. The weightings , w_m, for the each of the m channels of the MSLR patch, are computed by taking the values of the WorldView-2 Pan spectral response value at the centre of each of the MS channel response profiles, as defined by Full Width Half Maximum (HWFM). This is illustrated in Figure 5.9.

Figure 5.9: a) The WorldView-2 spectral responses, as per Figure 4.2. b) Computation of the weighting for channel 5, w_5, is depicted.

a)



b)



- WorldView 2 Pan Spectral Response Weighting, w_5, for Red Channel (5)
- Full Width at Half Maximum (FWHM) for Red Channel (5)

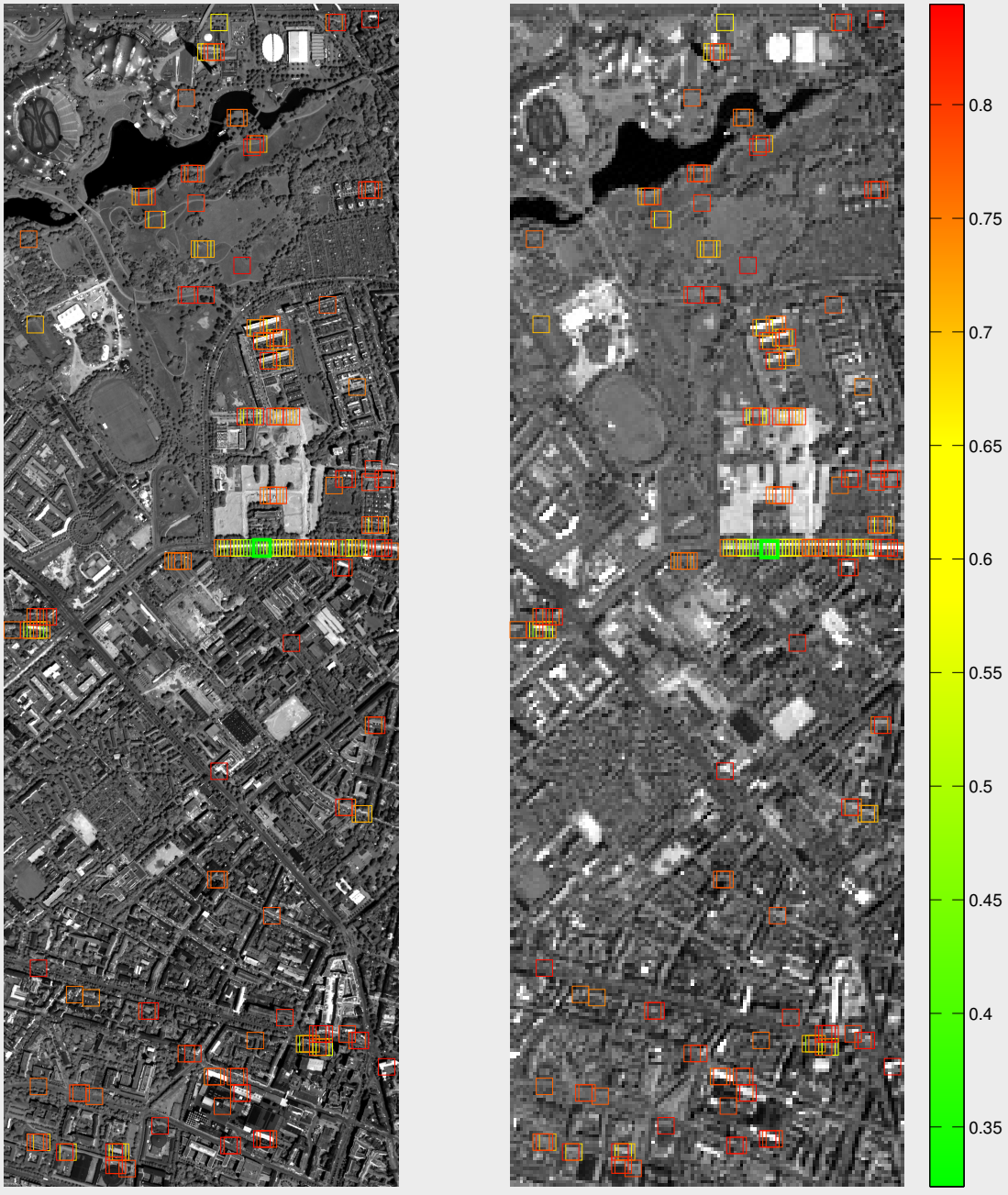
It is proposed that a dictionary for reconstructing the ith patch can be generated by comparing the euclidean difference between the ith approximate Pan patch, $\mathbf{x}^i_{l,approx}$, with the nth PanLR patch, $\mathbf{x}^n_l$:

$$\mu^{i,n}_{NL2} := \|\mathbf{x}^i_{l,approx} - \mathbf{x}^n_l\|_2 \tag{5.8}$$

The N_{DP} PanLR patches with the lowest euclidean norm difference, according to the metric, μ_{NL2} in Equation (5.8), are selected for the LR dictionary, $\mathbf{D}^n_l$. Corresponding PanHR patches are then selected for the HR dictionary, $\mathbf{D}^n_h$.

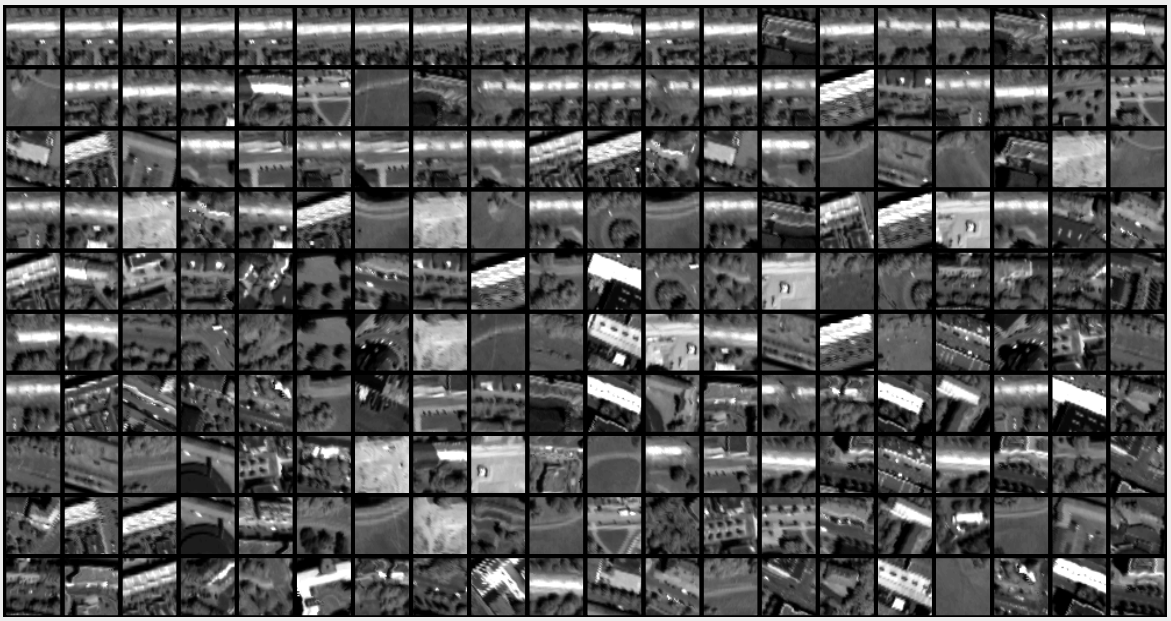
The NL2 dictionary location image is displayed in Figure 5.10.

Figure 5.10: The NL2 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the approximate Pan patch norm comparison, i.e μ_{NL2}.



The complete NL2 dictionary for $N_{DP} = 200$ is displayed in Figure 5.11.

Figure 5.11: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL2 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.4.2 NL2 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.12.

Figure 5.12: The performance of the NL2 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.

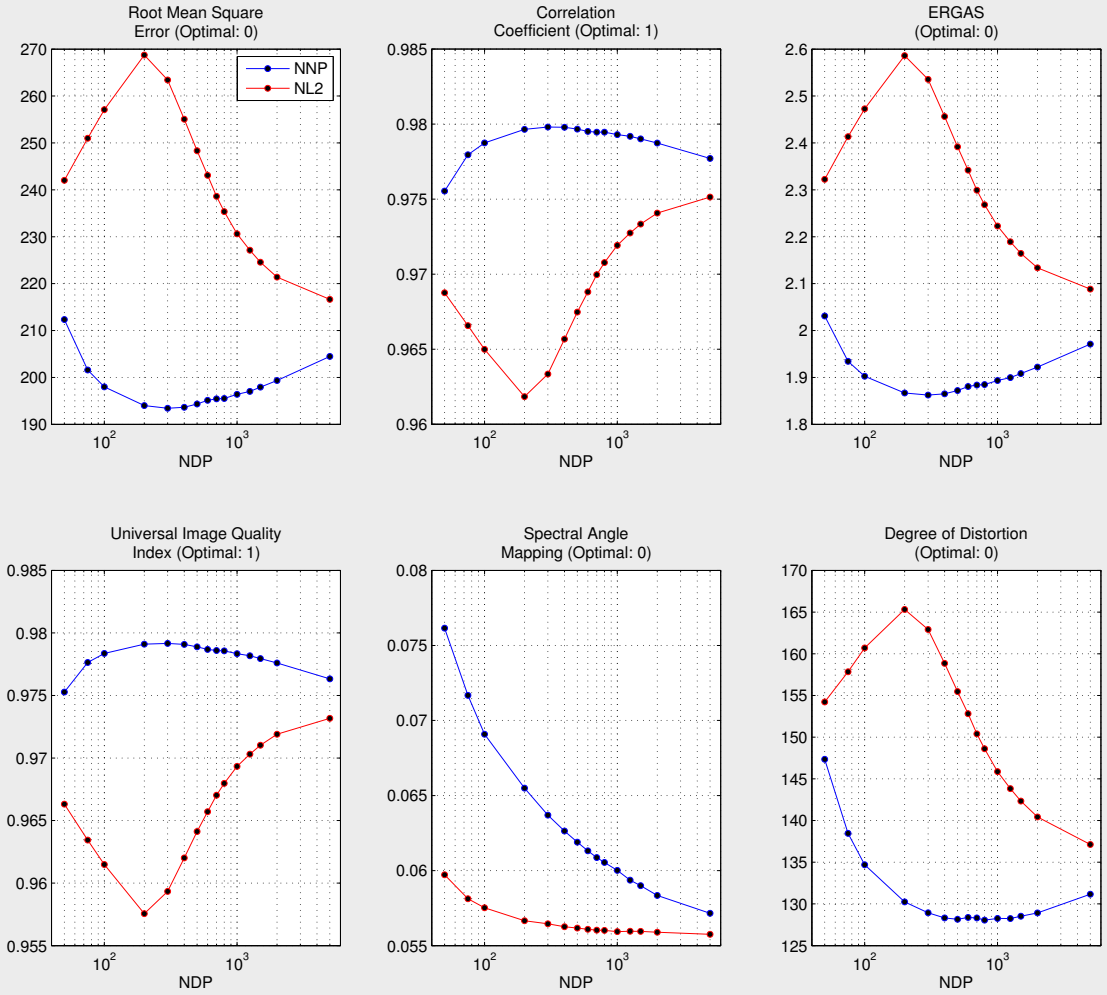


Figure 5.12 displays that NL2 dictionary selection also fails to improve on the benchmark NNP selection method in all metrics excluding the spectral angle mapping. However, also worthy of note, is that the NL2 dictionary selection method outperforms the NL1 method in all metrics, though it displays the same general trend. This provides some support that the

inclusion of the MSLR spectral information offers possible improvement in LR Pan similarity based dictionary selection.

5.5 Non-Local PanHR Norm Comparison Dictionary Selection (NL3)

5.5.1 Overview

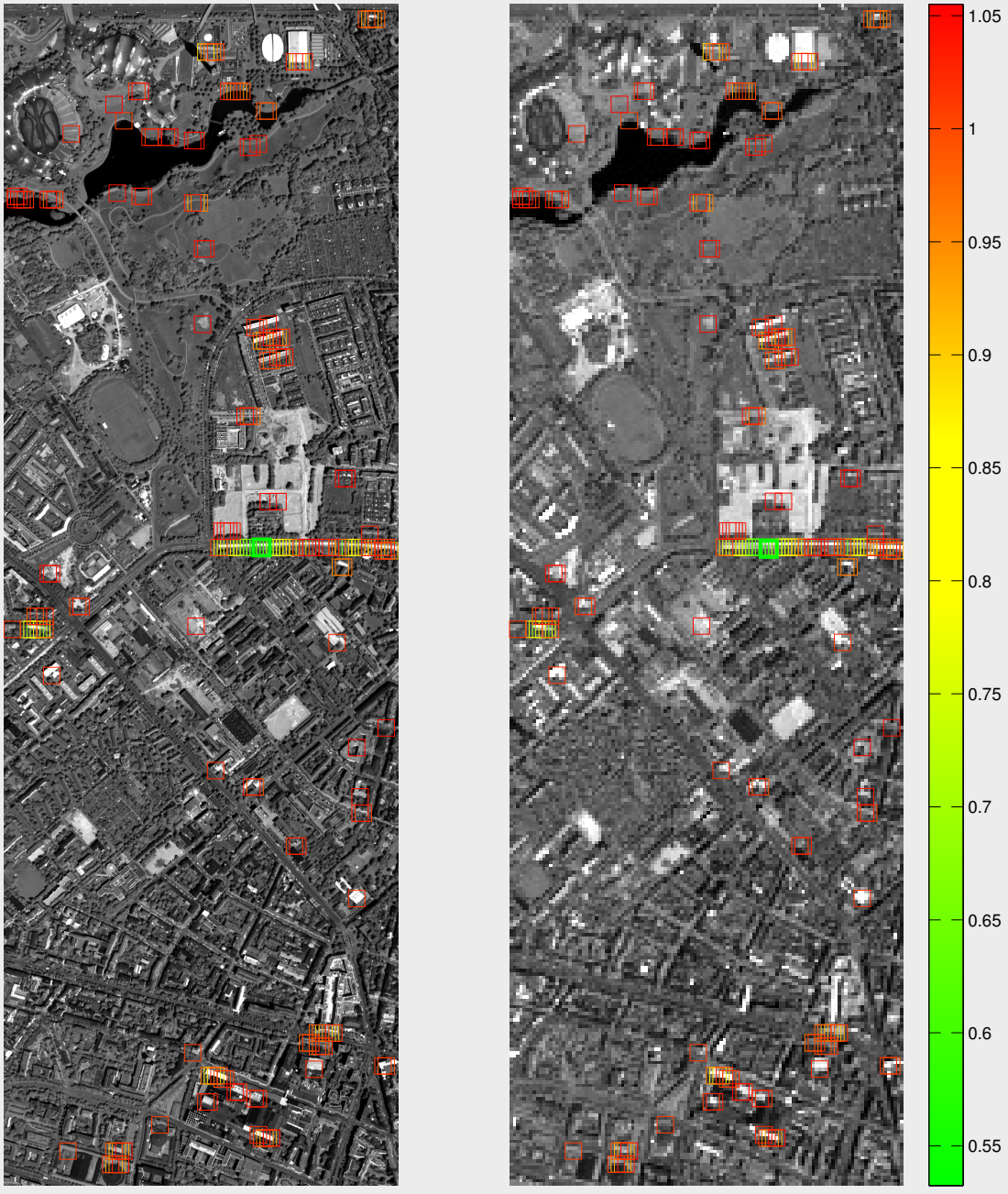
In an attempt to compensate for the loss of detail information involved in selecting dictionary patches on the basis of LR norm comparisons, the NL3 dictionary selection method employs HR norm comparison. Patches with the greatest similarity to the *ith* patch under reconstruction, in terms of euclidean norm difference between HR patches, are selected. Specifically, the euclidean norm of the difference between the *ith* vectorised PanHR patch under reconstruction, $\mathbf{x}_h^i$, and the *nth* patch, $\mathbf{x}_h^n$, from the complete set of N_P patches, is calculated:

$$\mu_{NL3}^{i,n} := \|\mathbf{x}_h^i - \mathbf{x}_h^n\|_2 \tag{5.9}$$

The N_{DP} most similar patches, according to this metric, correspond to the N_{DP} patches with the lowest euclidean norm difference in the HR system. These patch coordinates are subsequently taken and used to build the HR and LR coupled dictionaries.

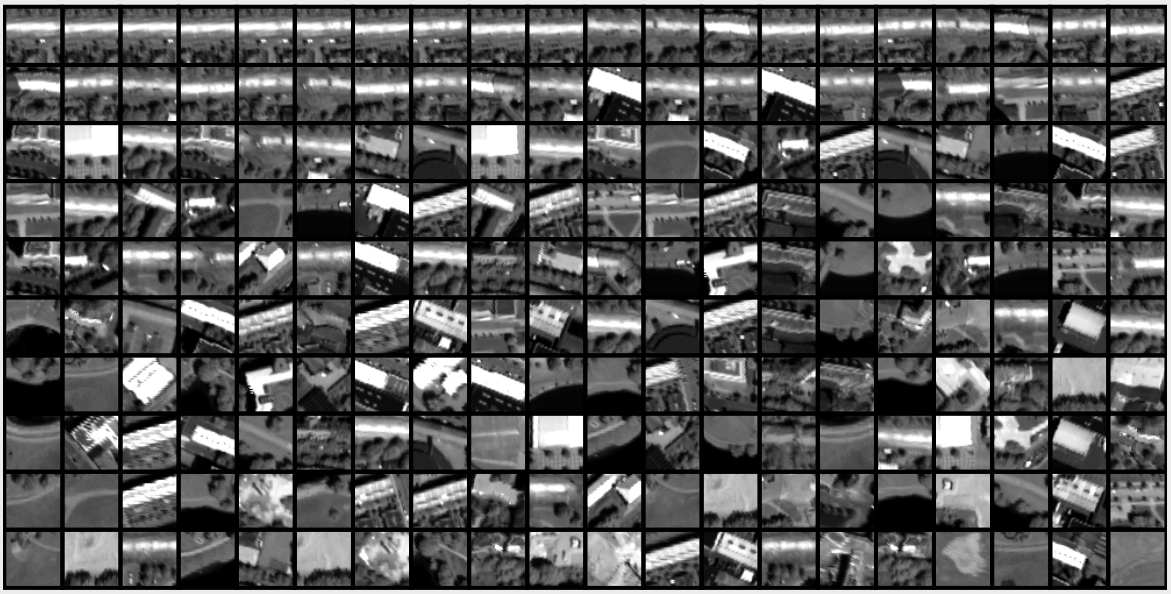
The NL3 dictionary location image is displayed in Figure 5.13.

Figure 5.13: The NL3 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the metric μ_{NL3}.



The complete NL3 dictionary for $N_{DP} = 200$ is displayed in Figure 5.14.

Figure 5.14: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL3 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.

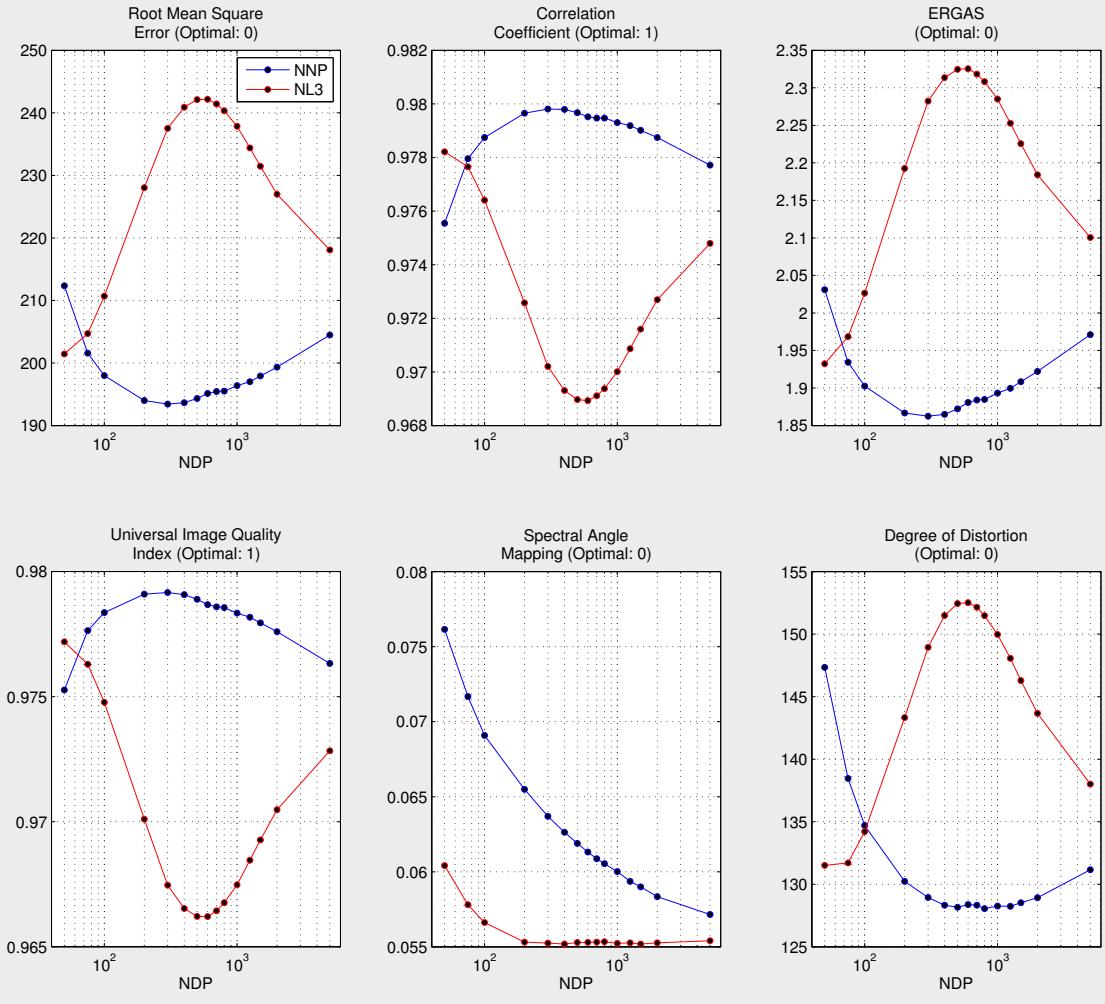


Note the prevalence of patches with red indicator boxes in Figure 5.13, depicting that the NL3 method rates a greater proportion of patches as dissimilar, when compared with the PanLR comparison methods, NL1 and NL2. This can be expected for this particular patch given that there are only a few street patches in the scene that are very similar to the current patch in high resolution. In low resolution, on the other hand, many other patches may also have a high similarity as discriminative details are lost due to their low resolution nature. Effectively, PanHR comparison methods can be expected to be more discriminative.

5.5.2 NL3 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.15.

Figure 5.15: The performance of the NL3 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



The same general trend for similarity based selection methods is observed, with a minimum performance peak or trough corresponding to peak performance in the benchmark NNP. It is important to note, however that NL3 outperforms both PanLR comparison methods, NL1 and NL2, and provides outstanding SAM metrics, well below the benchmark. Additionally, NL3 manages to outperform the benchmark for small dictionaries of ($N_{DP} \approx 50$).

5.6 Non-Local PanLR-PanHR Combined Rank Dictionary Selection (NL4)

The basis of the PanLR-PanHR Combined Rank Dictionary Selection (NL4) is the idea that selecting patches that rate highly in both HR and LR norm comparisons may strengthen the assumption that the sparse coefficients calculated for the LR system are valid for reconstructing the HR system.

In constructing coupled dictionaries for the *ith* patch under reconstruction, the following NL4 algorithm is computed:

NL4 Dictionary Selection Algorithm

---

1. Firstly, the N_{DP} patches are sorted on the basis of LR euclidean norm difference to the *ith* PanLR patch, as per the NL1 algorithm:

$$\mu_{NL1}^{i,n} := \|\mathbf{x}_l^i - \mathbf{x}_l^n\|_2 \tag{5.6}$$

   This LR patch sorting is converted to an ascending ranking, $r_l^i(n) \in \mathbb{N}_0$, beginning at 0 i.e the most similar patch in terms of the LR norm comparison would have

a ranking of $r_l^i(n = 1) = 0$. Note that this rank is in no way related to the mathematical definition of the rank of a matrix.

2. Following this, the N_{DP} patches are then sorted by means of HR euclidean difference to the *ith* PanHR patch, in accordance with the NL3 algorithm:

$$\mu_{NL3}^{i,n} = \|\mathbf{x}_h^i - \mathbf{x}_h^n\|_2 \tag{5.9}$$

In the same manner as step 1, the HR patch sorting is converted to an ascending ranking, $r_h^i(n) \in \mathbb{N}_0$, beginning at 0 i.e the most similar patch in terms of the HR norm comparison would have a ranking of $r_h^i(n = 1) = 0$.

3. A combined ranking, $r_c^i(n)$, is calculated for each patch in the total set of N_P patches, by summing the LR and HR rankings, $r_l^i(n)$ and $r_h^i(n)$ respectively:

$$r_c^i(n) = r_l^i(n) + r_h^i(n) \tag{5.10}$$

Note that by using a combined ranking, LR and HR euclidean difference comparisons are automatically given equal weighting. This is beneficial, as no additional consideration must be made to compensate for the different sizes of the LR and HR patches, and their corresponding difference in norm magnitudes.

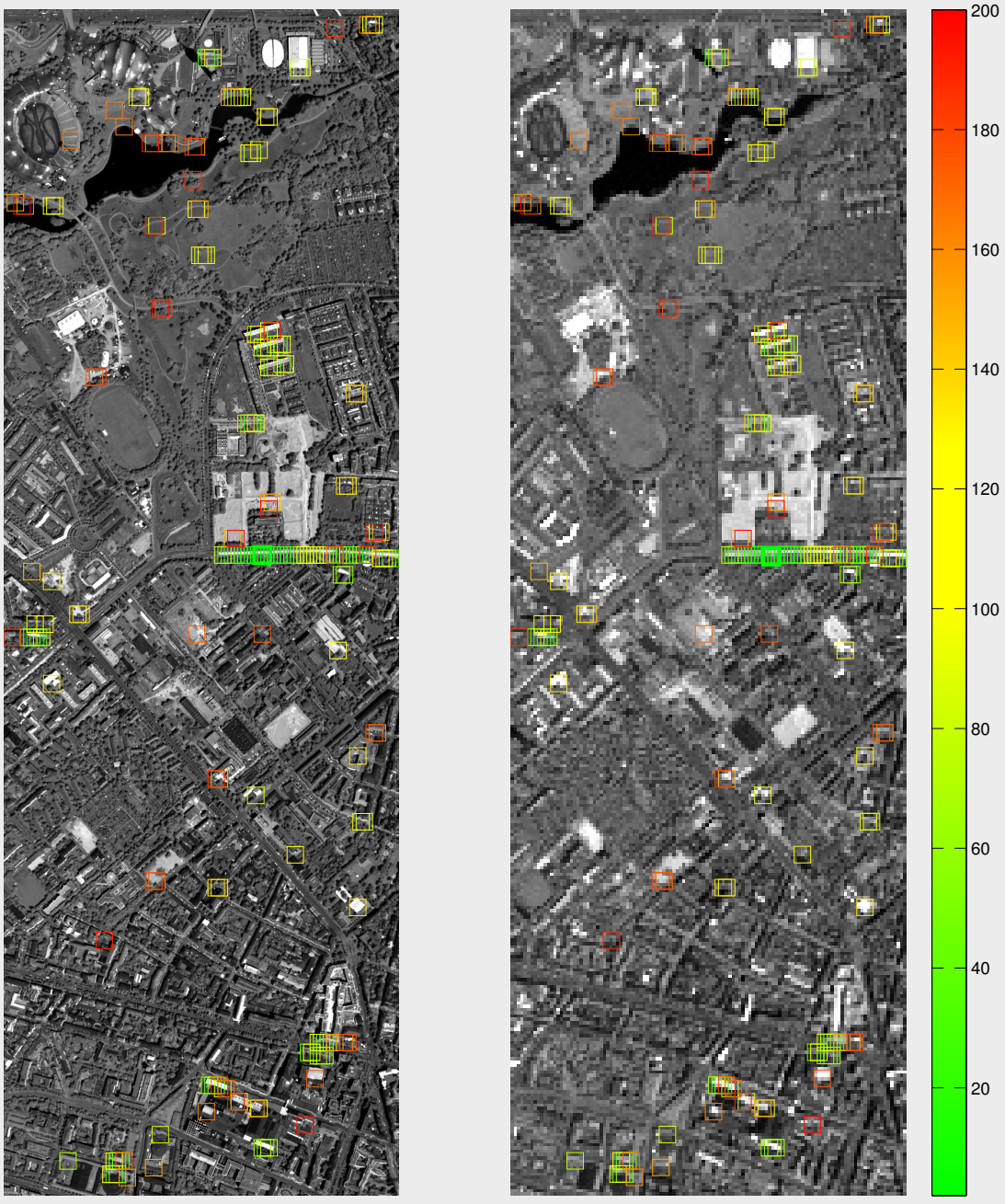
4. The set of N_{DP} patches with the lowest combined rankings, r_c^i, are taken to form the coupled LR and HR dictionaries. Thus we arrive at the following NL4 selection metric:

$$\mu_{NL4}^{i,n} = r_c^i(n) \tag{5.11}$$

Note that the additional storage matrices and sorting requirements inherent to the implementation of this method impact significantly on the computational efficiency.

The NL4 dictionary location image is displayed in Figure 5.16.

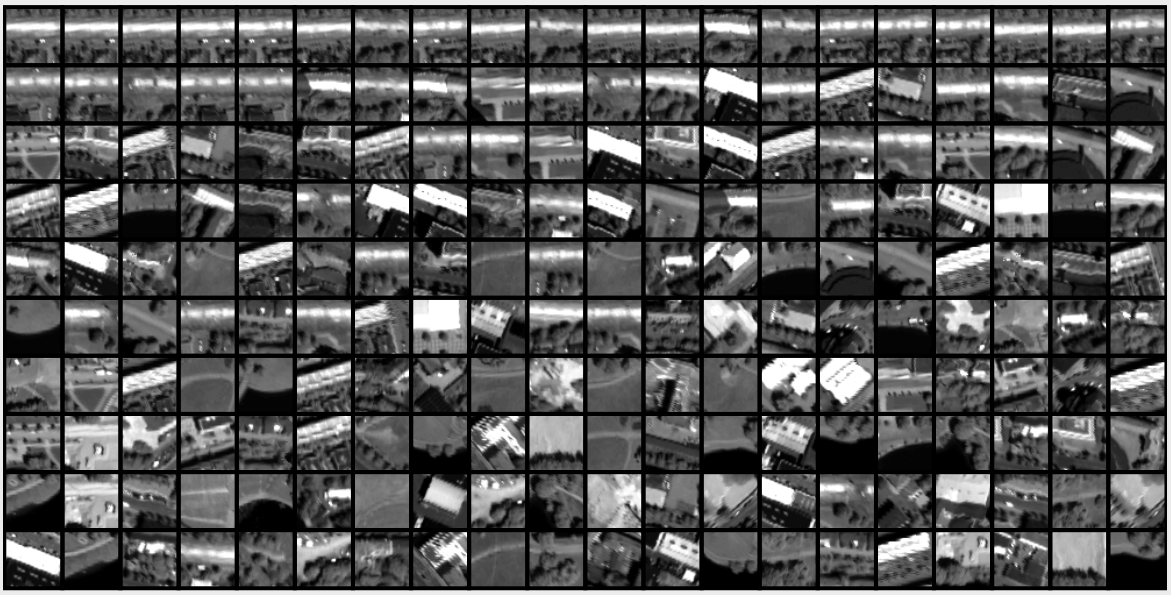
Figure 5.16: The NL4 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the combined PanHR-PanLR ranking i.e. μ_{NL4}.



The high combined ranking of road patches is evident through the clustering of green patch indicators the road surrounding the example patch.

The complete NL4 dictionary for $N_{DP} = 200$ is displayed in Figure 5.17.

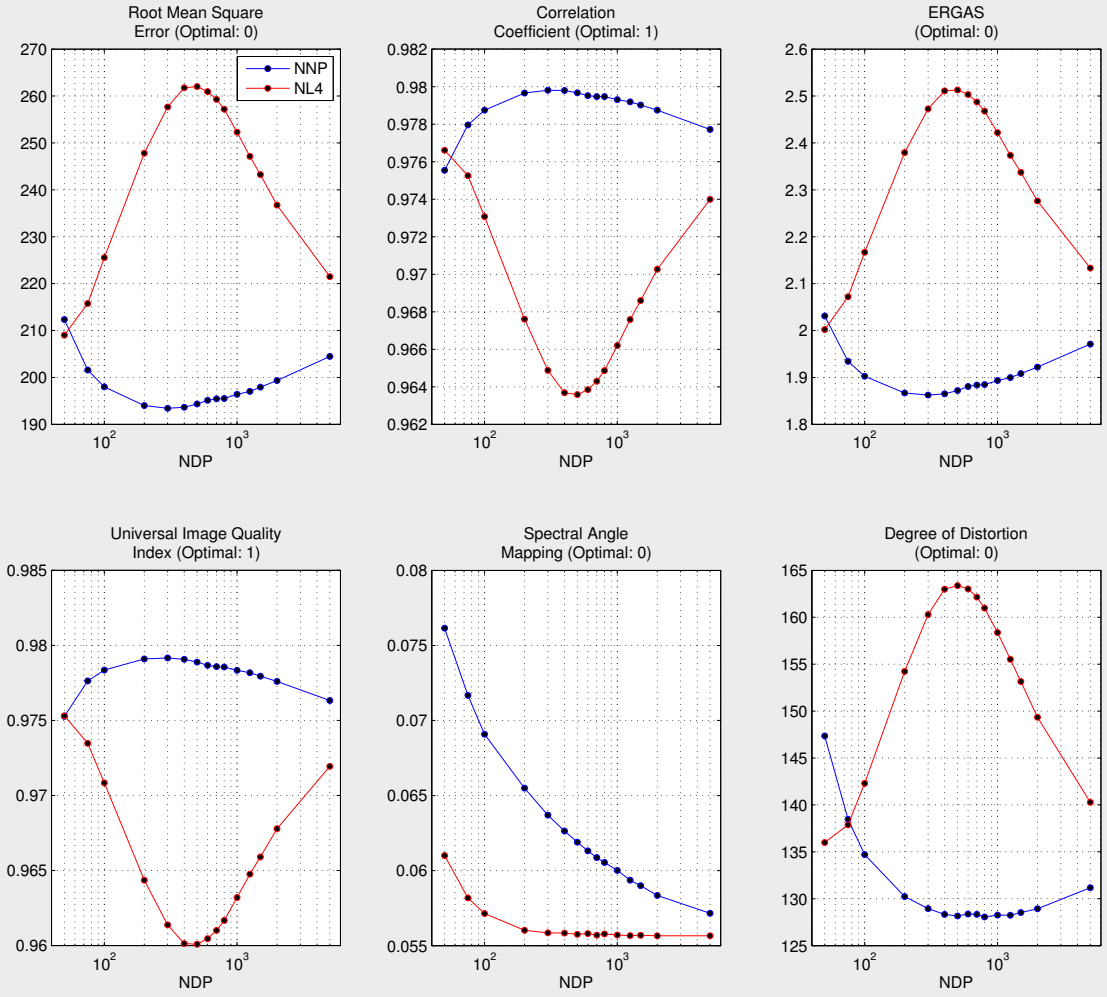
Figure 5.17: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL4 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.6.1 NL4 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.18.

Figure 5.18: The performance of the NL4 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



Interestingly the performance of the NL4 selection method falls between the range of the PanLR selection method, NL1 and the PanHR selection method, NL3. Since, fundamentally, NL4 is a combination of these HR and LR norm comparison methods, the intermediate performance is a reasonable result. This demonstrates, however, that choosing a dictionary

based on strengthening the connection between HR and LR systems does not necessarily improve performance of J-SparseFI.

5.7 Non-Local PanHR Absolute Correlation Dictionary Selection (NL5)

The Non-Local PanHR Absolute Correlation Dictionary selection method (NL5) attempts to exploit coupled dictionaries containing patches with both positive and negatively correlated information. More specifically, given the *ith* vectorised PanHR patch corresponding to the *ith* patch under reconstruction, $\mathbf{x}_h^i$, and the *nth* PanHR patch under consideration, $\mathbf{x}_h^n$, the correlation coefficient can be calculated as:

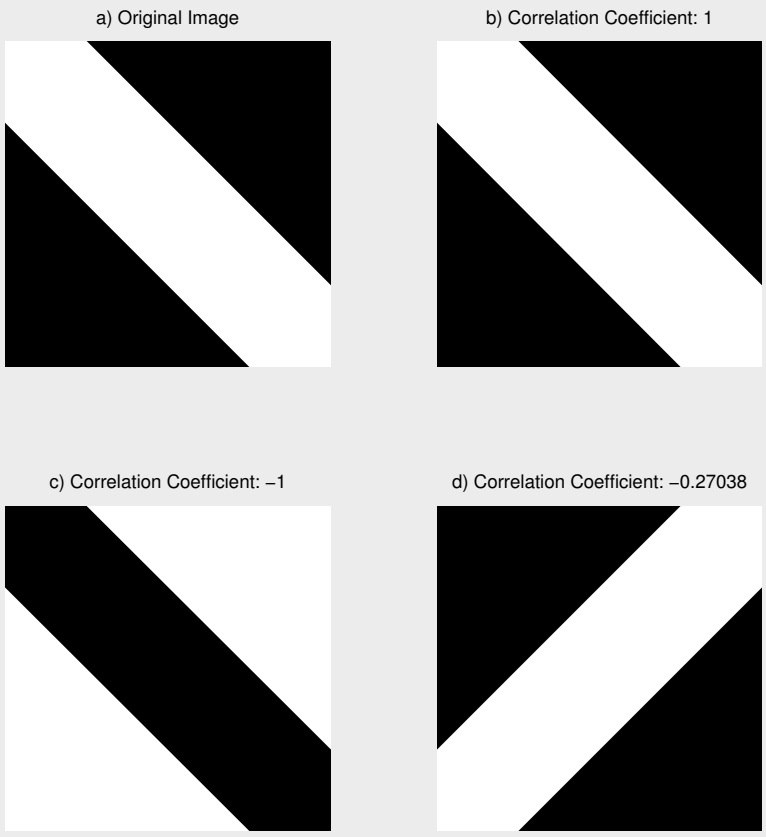
$$\rho^{i,n} = \frac{\sum_{p}^{P_h} \left[\left(\mathbf{x}_h^i(p) - \bar{\mathbf{x}}_h^i \right) \cdot \left(\mathbf{x}_h^n(p) - \bar{\mathbf{x}}_h^n \right) \right]}{\sqrt{\left[\sum_{p}^{P_h} \left(\mathbf{x}_h^i(p) - \bar{\mathbf{x}}_h^i \right)^2 \right] \cdot \left[\sum_{p}^{P_h} \left(\mathbf{x}_h^n(p) - \bar{\mathbf{x}}_h^n \right)^2 \right]}} \tag{5.12}$$

Where $\bar{\mathbf{x}}_h^i$ and $\bar{\mathbf{x}}_h^n$ are the mean values of the *ith* and *nth* patches respectively. The correlation coefficient is outlined visually in Figure 5.19. The absolute correlation of patches is thus used as a selection metric:

$$\mu_{NL5}^{i,n} := |\rho^{i,n}| \tag{5.13}$$

Remark 5.2. *In our specific image fusion problem, we are actually subtracting the mean value and normalising the patches prior to patch selection, in accordance with Equation* (5.3) *and Equation* (5.4) *in Section 5.1. Hence the two norms in Equation* (5.12) *are both equal to one.*

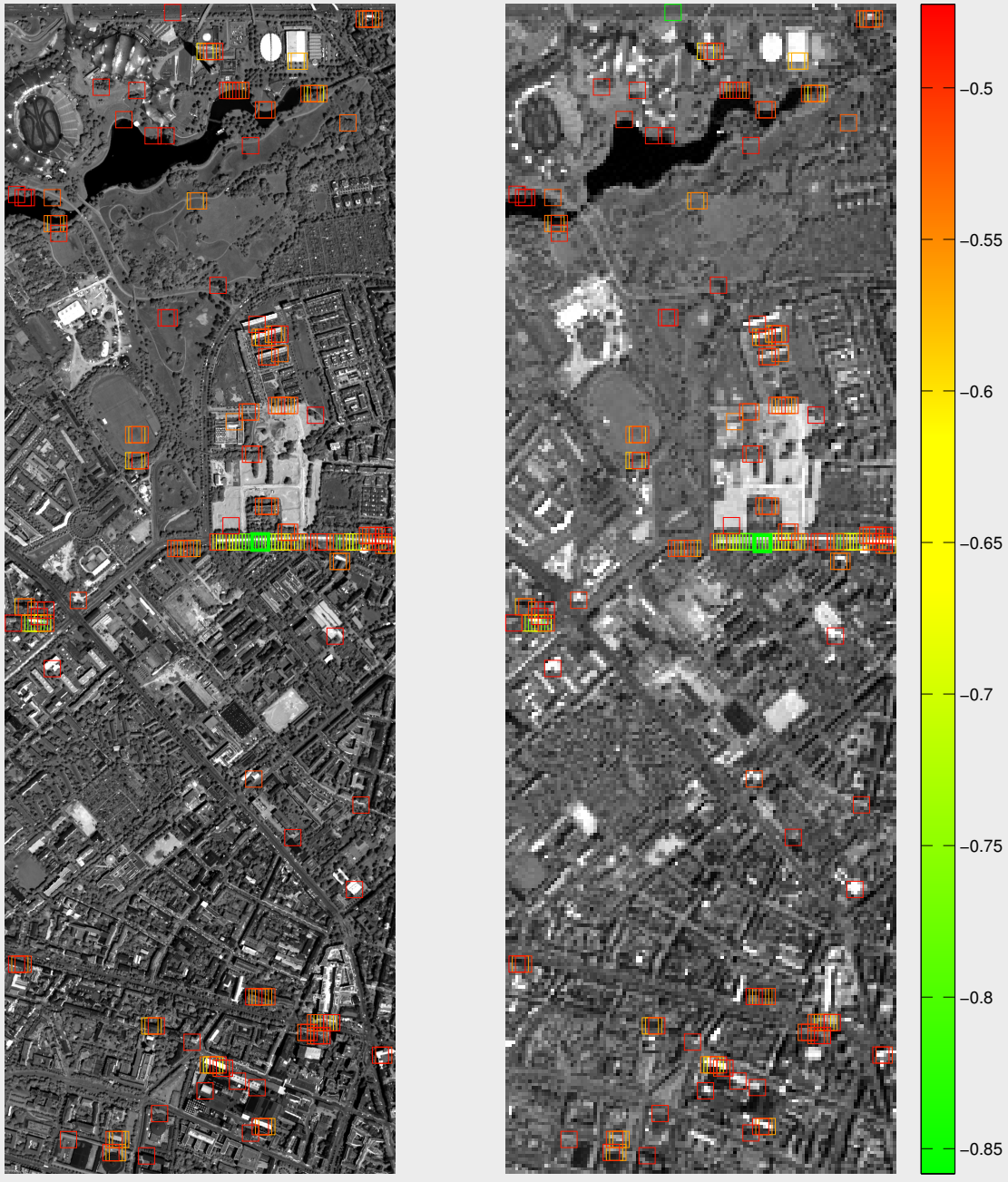
Figure 5.19: This diagram depicts correlation based similarity. An image, b), that is an exact copy of the original image, a) has a correlation coefficient of 1. An image, c), that is an inverse of the original image a) has a correlation coefficient of -1. A flipped version of the original image, d), has a correlation coefficient relatively close to 0.



Note that since Equation (5.13) evaluates patches on the basis of *absolute* correlation, negatively correlated patches will also be included, providing the dictionaries with additional geometrically correlated patches, in comparison with norm based methods, such as NL3. Also note that the flipped image d) in Figure 5.19 demonstrates that patches containing different geometry will rank low in their selection metric, despite possible visual or scene similarity.

The NL5 dictionary location image is displayed in Figure 5.20.

Figure 5.20: The NL5 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of absolute PanHR correlation i.e μ_{NL5}.



The operation of the absolute correlation coefficient is apparent in Figure 5.21.

Figure 5.21: The selection of corresponding vegetation patches, with a high negative correlation to the example road patch under reconstruction, is displayed a) HR and b) LR Pan images.

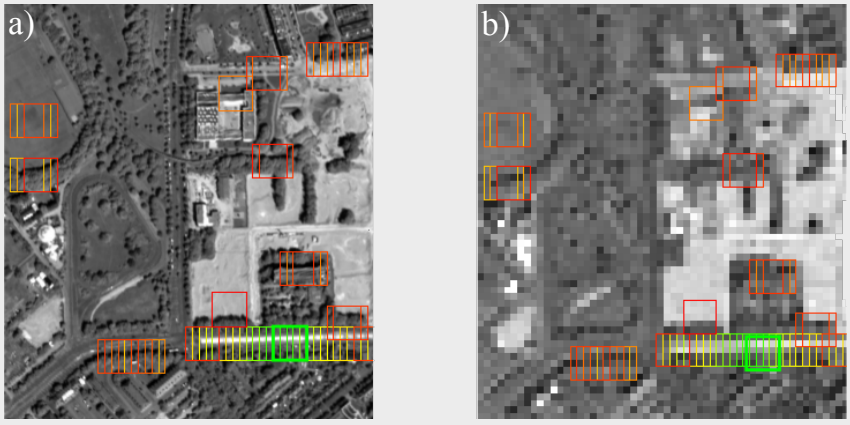
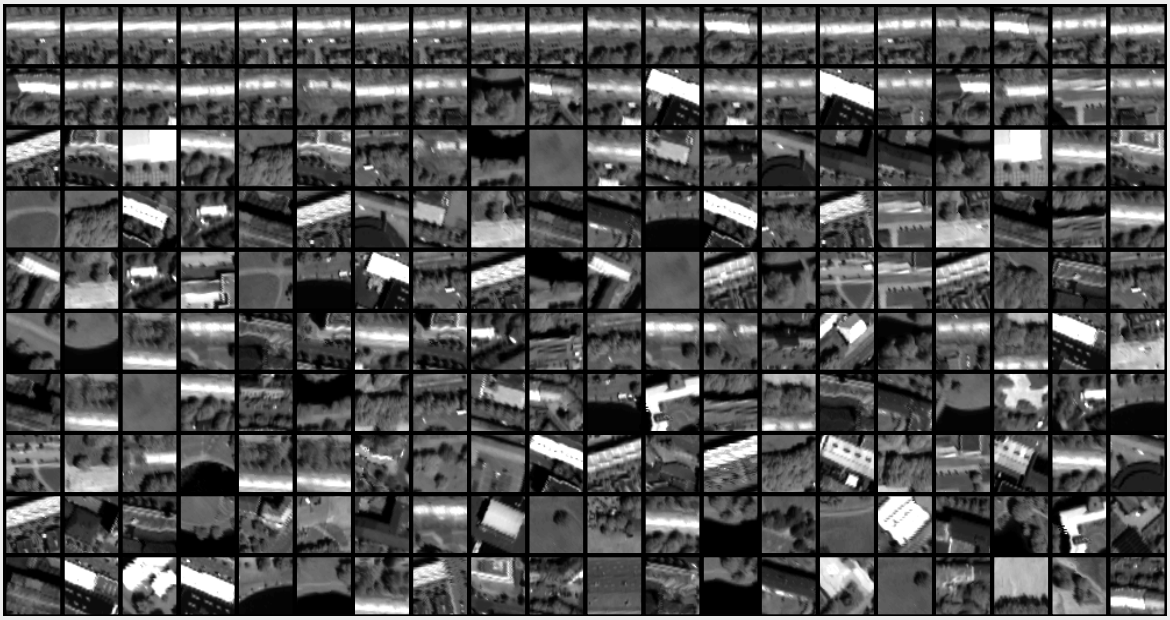


Figure 5.21 Depicts the selection of vegetation patches, with corresponding horizontal textures, but an inverse intensity. The horizontal line of trees casts a shadow, which is comparably located with respect to the road position in the patch undergoing reconstruction. This provides verification that the selection method is operating as intended.

The complete NL5 dictionary for $N_{DP} = 200$ is displayed in Figure 5.22.

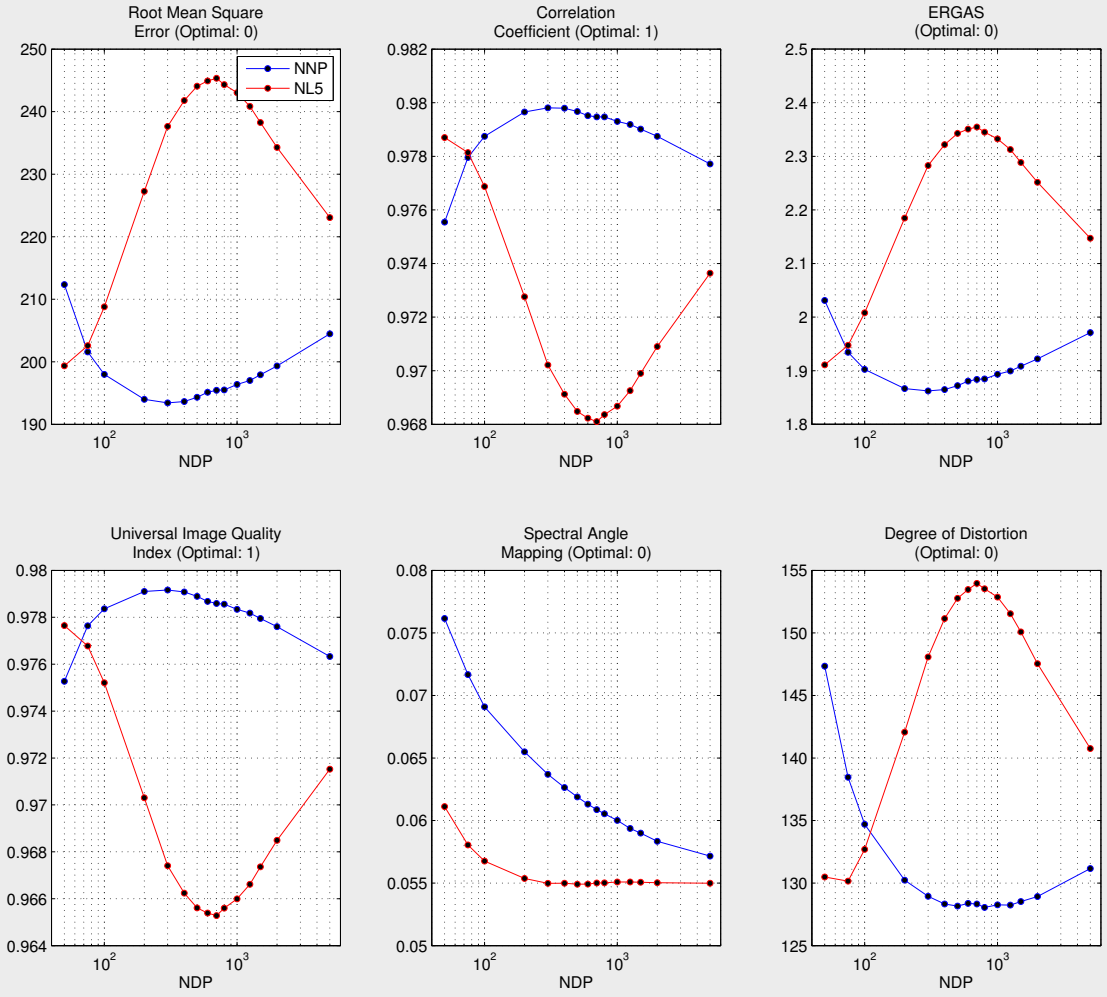
Figure 5.22: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL5 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.7.1 NL5 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.23.

Figure 5.23: The performance of the NL5 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



The performance of the NL5 dictionary selection method matches the performance of the NL3 selection method, with the same considerable improvement in SAM.

5.8 Non-Local PanHR Positive Correlation Dictionary Selection (NL6)

In order to properly evaluate correlation based methods, positive correlation matching is used to form the coupled dictionaries for the NL6 method. Comparison and categorisation with regards to the absolute correlation comparison method, NL5, is then possible.

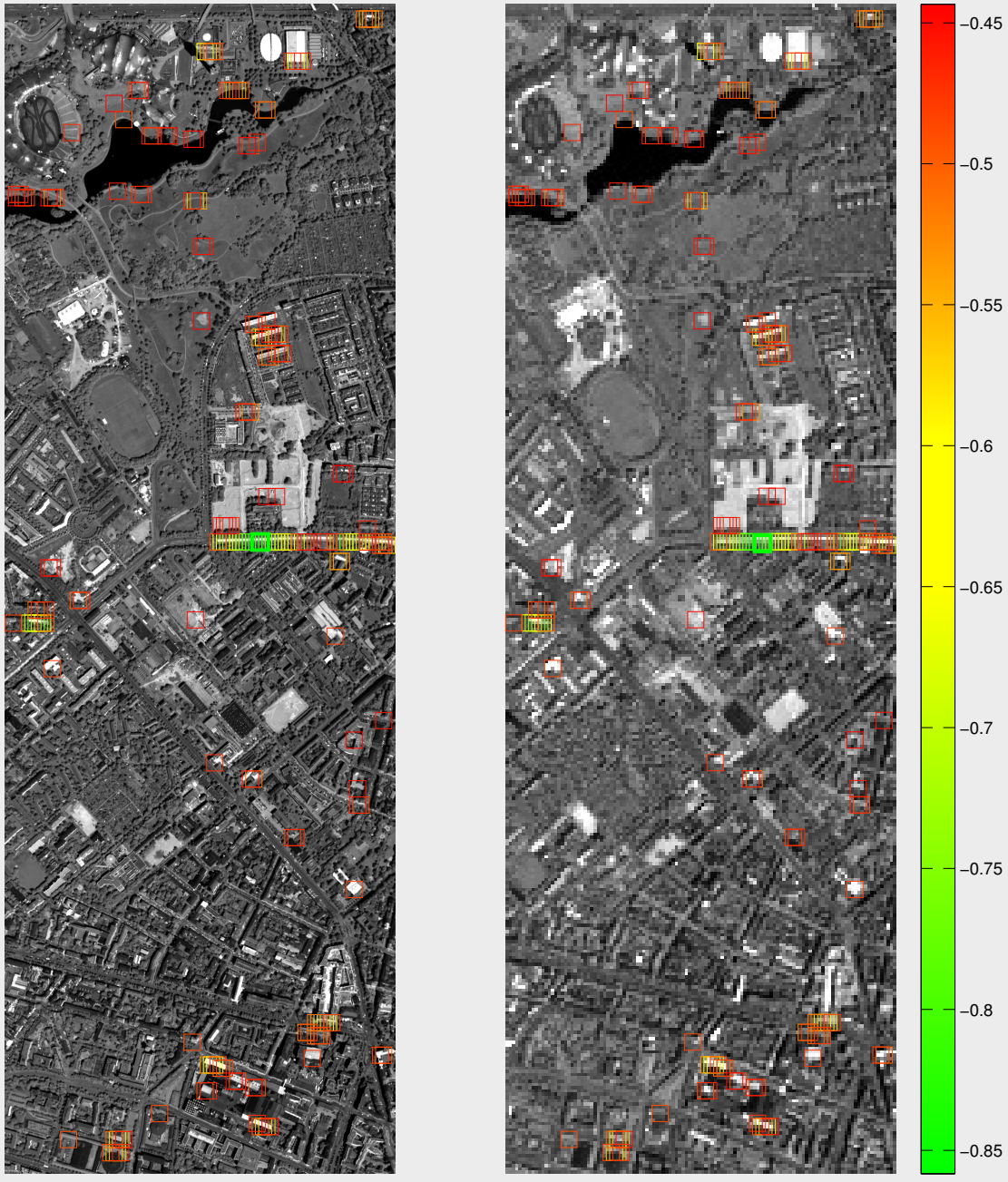
Correlation is calculated in accordance with Equation (5.12). However, in contrast to the measure $\mu_{NL5}^{i,n}$ defined in Equation (5.13), the absolute correlation is not taken:

$$\mu_{NL6}^{i,n} := \rho^{i,n} \tag{5.14}$$

Thus the coupled dictionaries are composed of patches that positively correlate to the current patch under reconstruction.

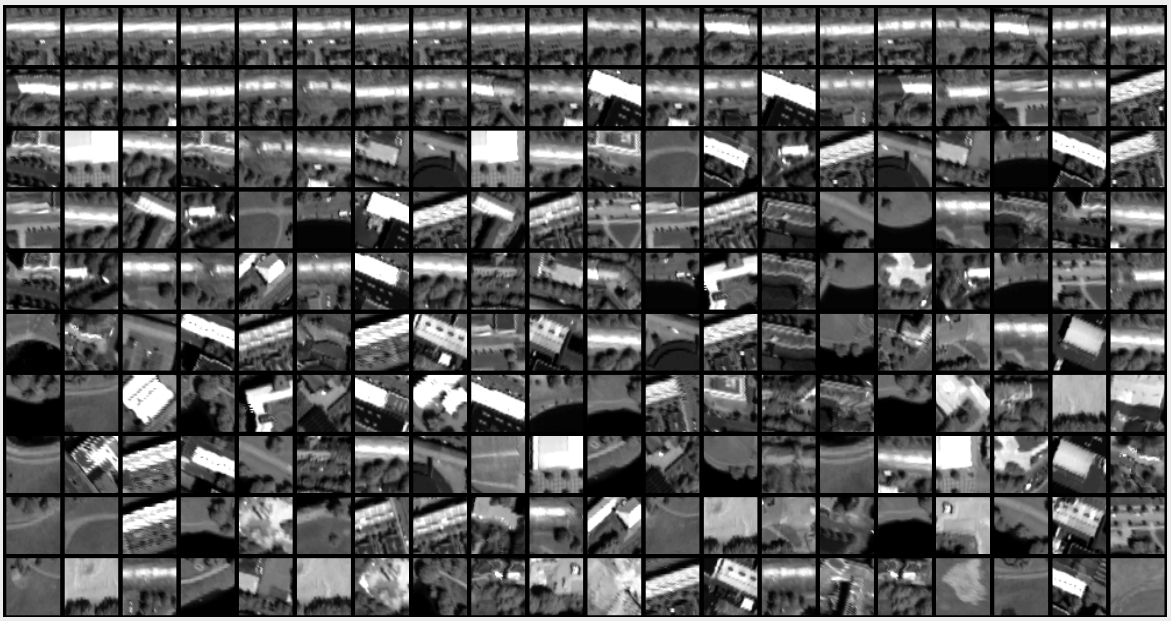
The NL6 dictionary location image is displayed in Figure 5.24.

Figure 5.24: The NL6 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of positive PanHR correlation i.e. μ_{NL6}.



The complete NL6 dictionary for $N_{DP} = 200$ is displayed in Figure 5.25.

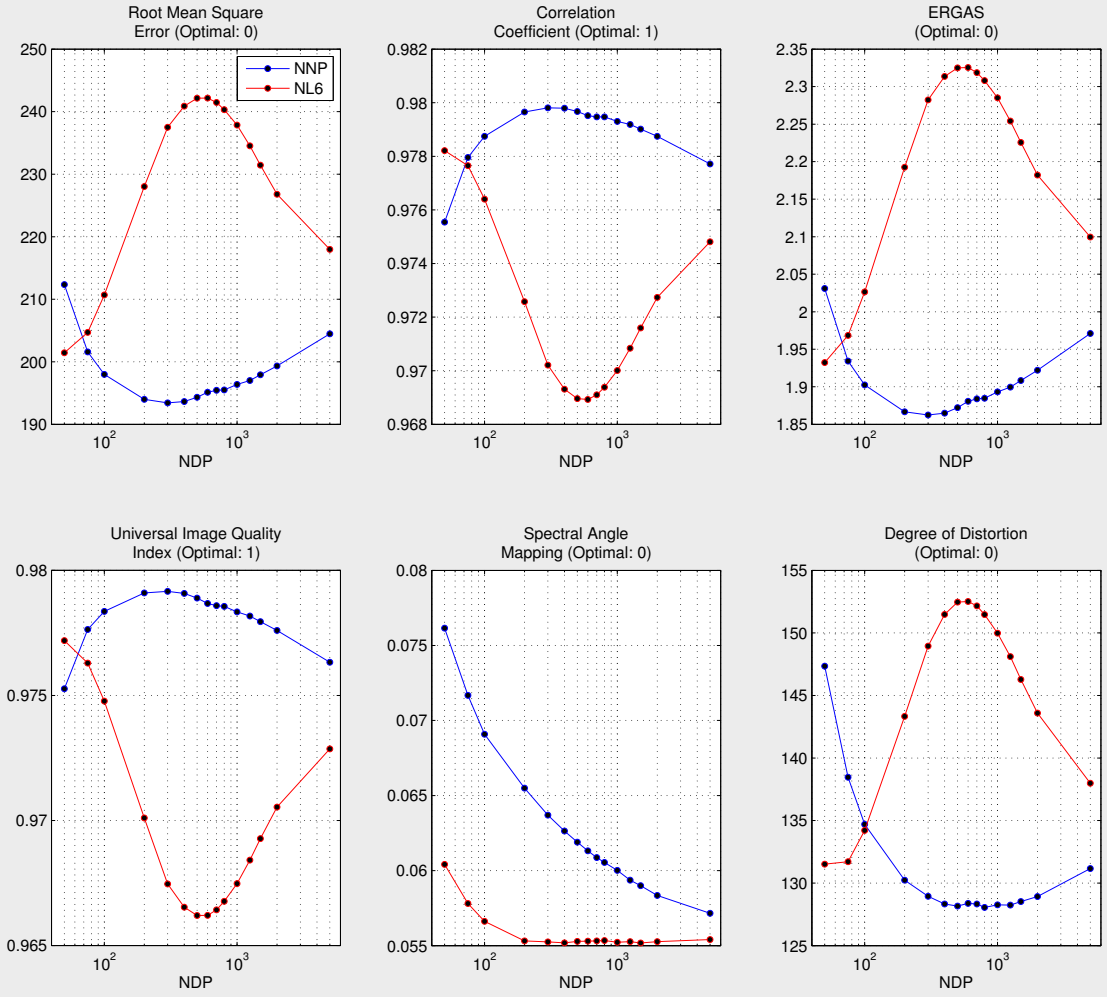
Figure 5.25: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL6 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.8.1 NL6 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.26.

Figure 5.26: The performance of the NL1 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



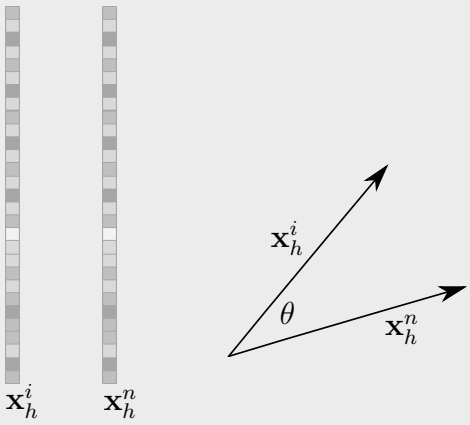
The performance of the NL6 dictionary selection method matches the performance of the NL3 selection method, with the characteristic considerable improvement in SAM. This is because, the patches get normalized and their means get subtracted prior to any patch selection method. Therefore, in this particular J-SparseFI image fusion configuration, the

two measures μ_{NL3} and μ_{NL6} are equivalent. Hence, the identical performance of NL3 and NL6. Also observe that, compared with NL5, there is not a significant difference in performance. This indicates that, at least for the data and configuration used in this thesis, including negatively correlated patches in the dictionary does not significantly change the overall performance of J-SparseFI.

5.9 Non-Local PanHR Patch Angle Mapping Dictionary Selection (NL7)

The Non-Local PanHR Patch Angle Mapping Dictionary Selection Method (NL7), is another HR similarity based metric. This method, inspired by the Spectral Angle Mapping metric, treats the vectorised PanHR patches as vectors in HR patch size P_h dimensional space. The similarity measurement is then taken to be the angle between the vectorised patches, as depicted in Figure 5.27.

Figure 5.27: The vectorised patches $\mathbf{x}_h^i$, corresponding to the patch undergoing reconstruction, and $\mathbf{x}_h^n$, the nth dictionary patch, can be represented as vectors in P_h dimensional space. The angle between such vectors thus provides a measure of patch similarity.



Specifically given the *ith* vectorised PanHR patch corresponding to the *ith* patch under reconstruction, $\mathbf{x}_h^i$, and the *nth* PanHR patch under consideration, $\mathbf{x}_h^n$, the Patch Angle Mapping (PAM) metric, $\mu_{NL7}$is calculated as:

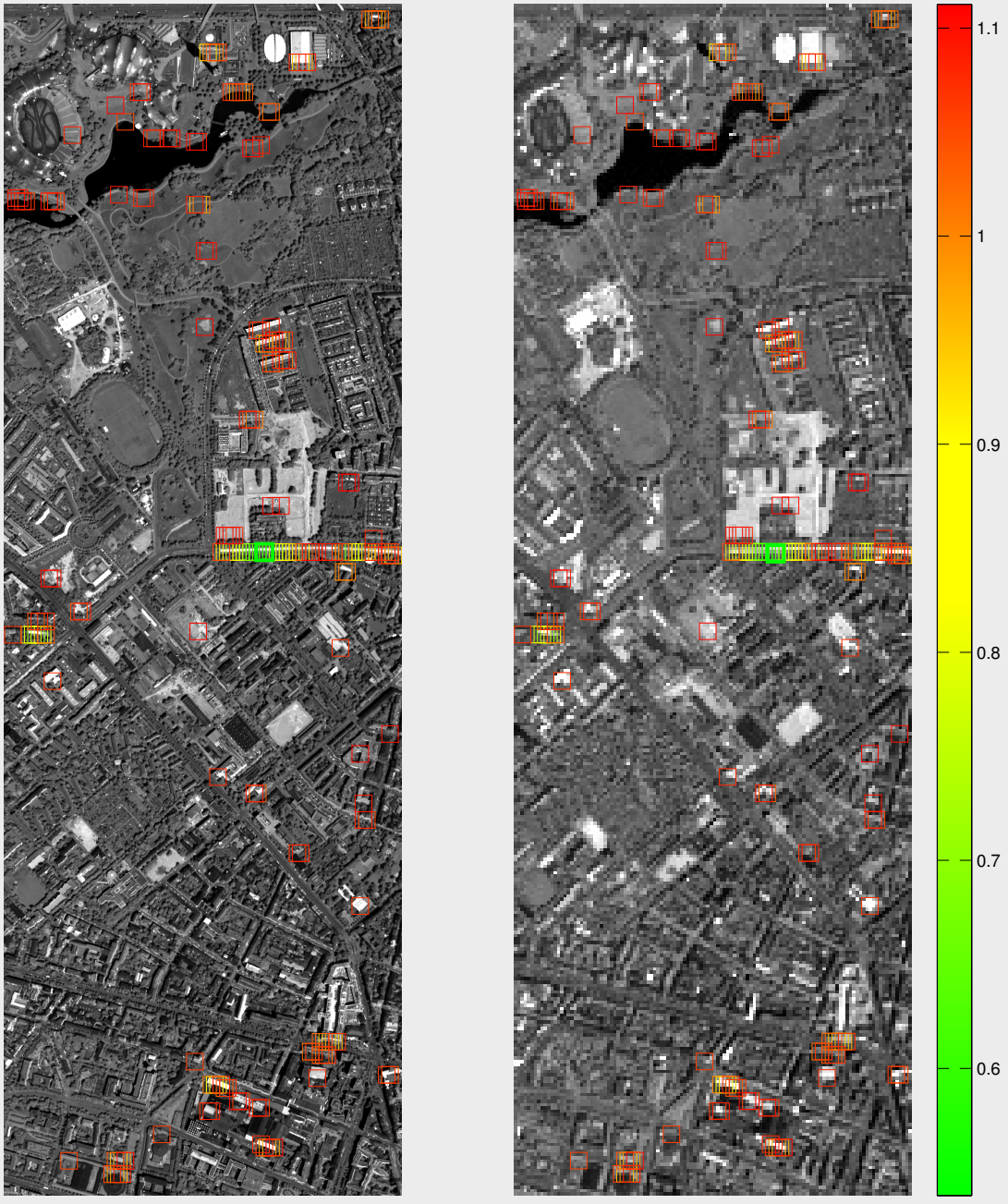
$$\mu_{NL7}^{i,n} := \theta = \arccos\left(\frac{(\mathbf{x}_h^i)^T\mathbf{x}_h^n}{\|\mathbf{x}_h^i\|_2\|\mathbf{x}_h^n\|_2}\right) \tag{5.15}$$

Importantly, this method is independent of the patch intensity, since this only represents an increase in vector magnitude. The relative intensities of pixels is the determining factor in the angular displacement of the patches in vector space, making this method a robust selection method for varying scene lighting. This method is expected to select very similar dictionaries to NL3 and NL6 methods.

Remark 5.3. *Observe that since arccos:* $[-1, 1] \to [0, \pi]$ *is a strictly monotonically decreasing function, the* N_{DP} *patches with the lowest patch angle,* μ_{NL7}*, will be the same as the* N_{DP} *patches with the highest correlation,* μ_{NL6}*. Therefore, the NL7 results are identical to the results achieved using NL6 and NL3.*

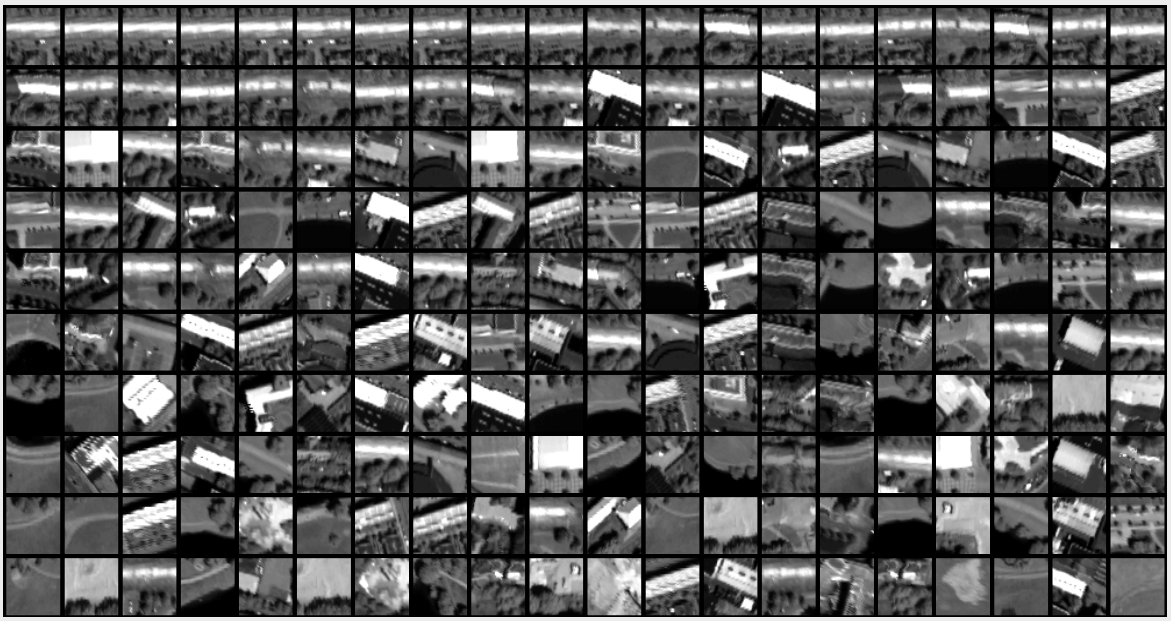
The NL7 dictionary location image is displayed in Figure 5.28.

Figure 5.28: The NL7 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of the PanHR Patch Angle Mapper i.e. μ_{NL7}.



The complete NL7 dictionary for $N_{DP} = 200$ is displayed in Figure 5.29.

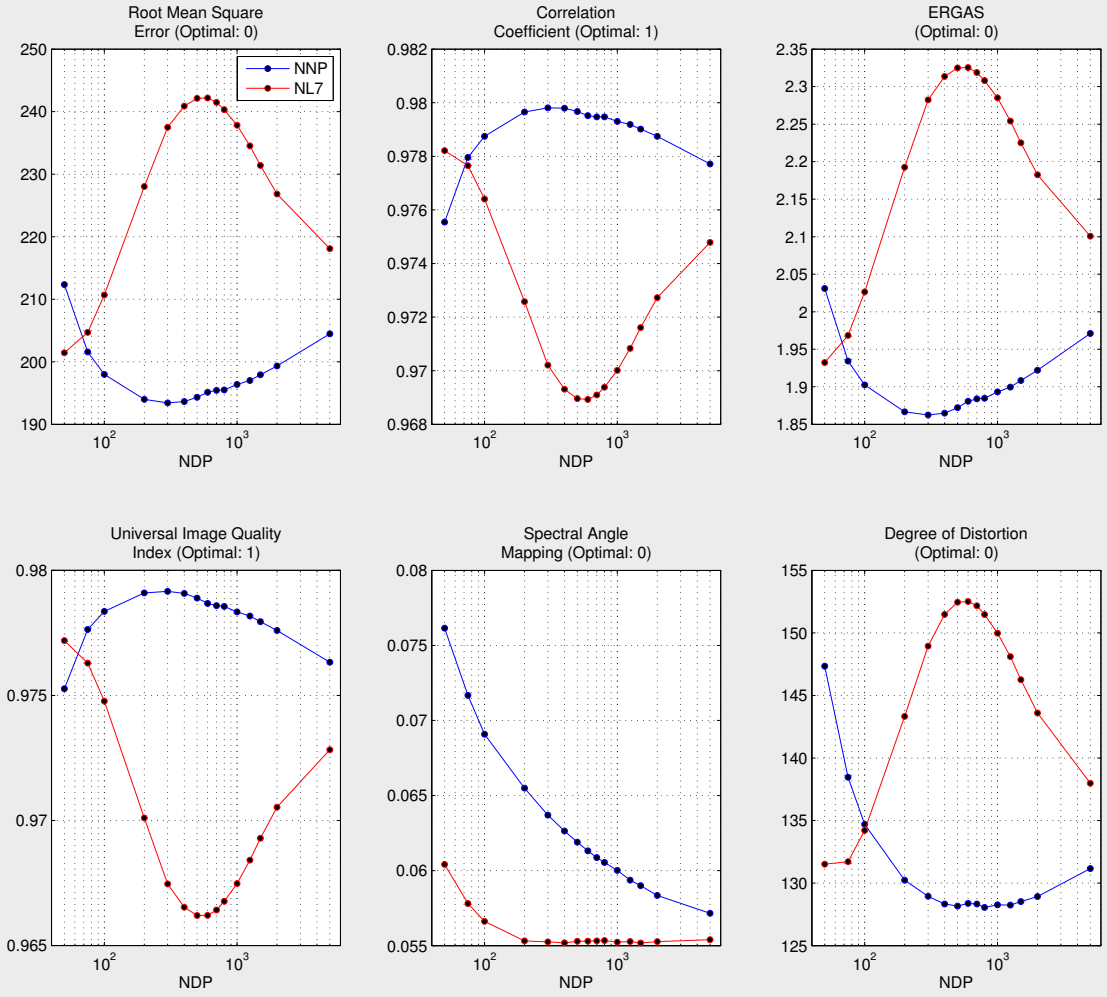
Figure 5.29: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL7 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.9.1 NL7 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.30.

Figure 5.30: The performance of the NL7 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



The performance of the NL7 dictionary selection method matches the performance of HR similarity based NL3 and NL6 dictionary selection methods, with the characteristic considerable improvement in SAM.

5.10 Non-Local PanHR Anti-Correlation Dictionary Selection (NL8)

A different approach was undertaken for the Non-Local PanHR Anti-Correlation Dictionary Selection method (NL8). Considering the inferior spatial performance of the similarity based selection methods, coupled dictionaries were selected on the basis of PanHR anti-correlation to the patch currently under reconstruction.

The NL8 uses the same correlation coefficient metric as detailed for the absolute correlation selection method, NL5, as detailed in Equation (5.13), however patches are selected on the basis of lowest absolute correlation i.e patches with an absolute correlation coefficient that approaches zero.

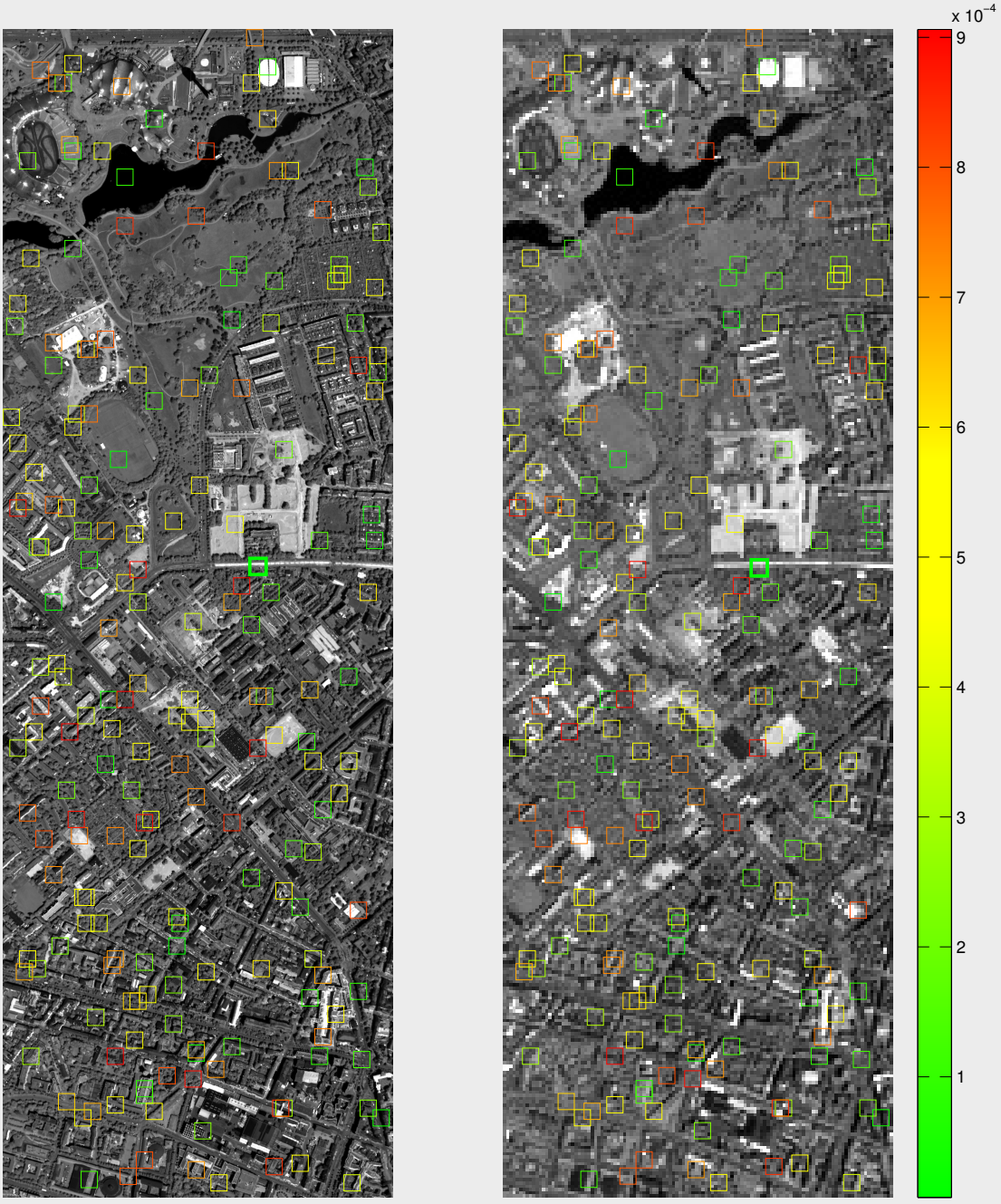
Hence, the patches that return the smallest values measured by Equation (5.16) are selected:

$$\mu_{NL8}^{i,n} := \mu_{NL5}^{i,n} \tag{5.16}$$

Note that the first element in the dictionary was taken to correspond to the current patch under reconstruction i.e. the *ith* PanHR and PanLR patches, corresponding to the *ith* patch under reconstruction, are respectively included as the first element in the HR and LR dictionaries respectively. Following this the $N_{DP} - 1$ patches with the lowest PanHR correlation according to Equation (5.16) to the *ith* PanHR patch, will be selected to form the coupled HR and LR dictionaries.

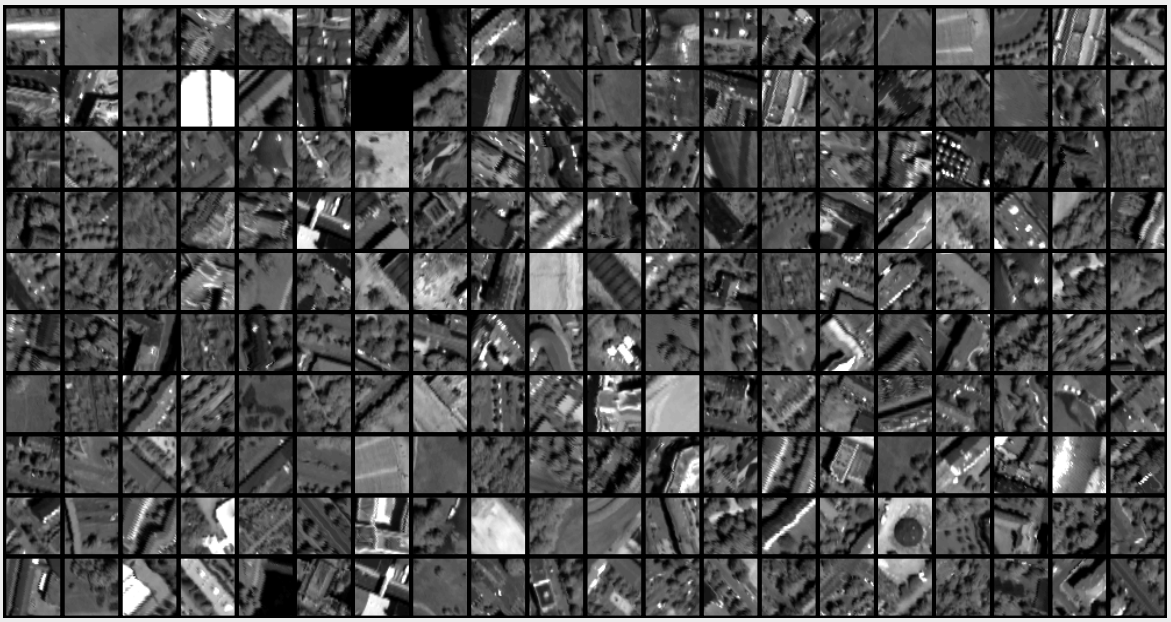
The NL8 dictionary location image is displayed in Figure 5.31.

Figure 5.31: The NL8 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the patch ranking in terms of PanHR patch anti-correlation i.e. μ_{NL8}.



The anti-correlation selection is evident in Figure 5.31, where none of the surrounding road patches have been selected. In fact there is no evidence of any patches bearing horizontal geometry resembling the example road patch, as depicted in the complete NL8 dictionary. The complete NL8 dictionary for $N_{DP} = 200$ is displayed in Figure 5.32.

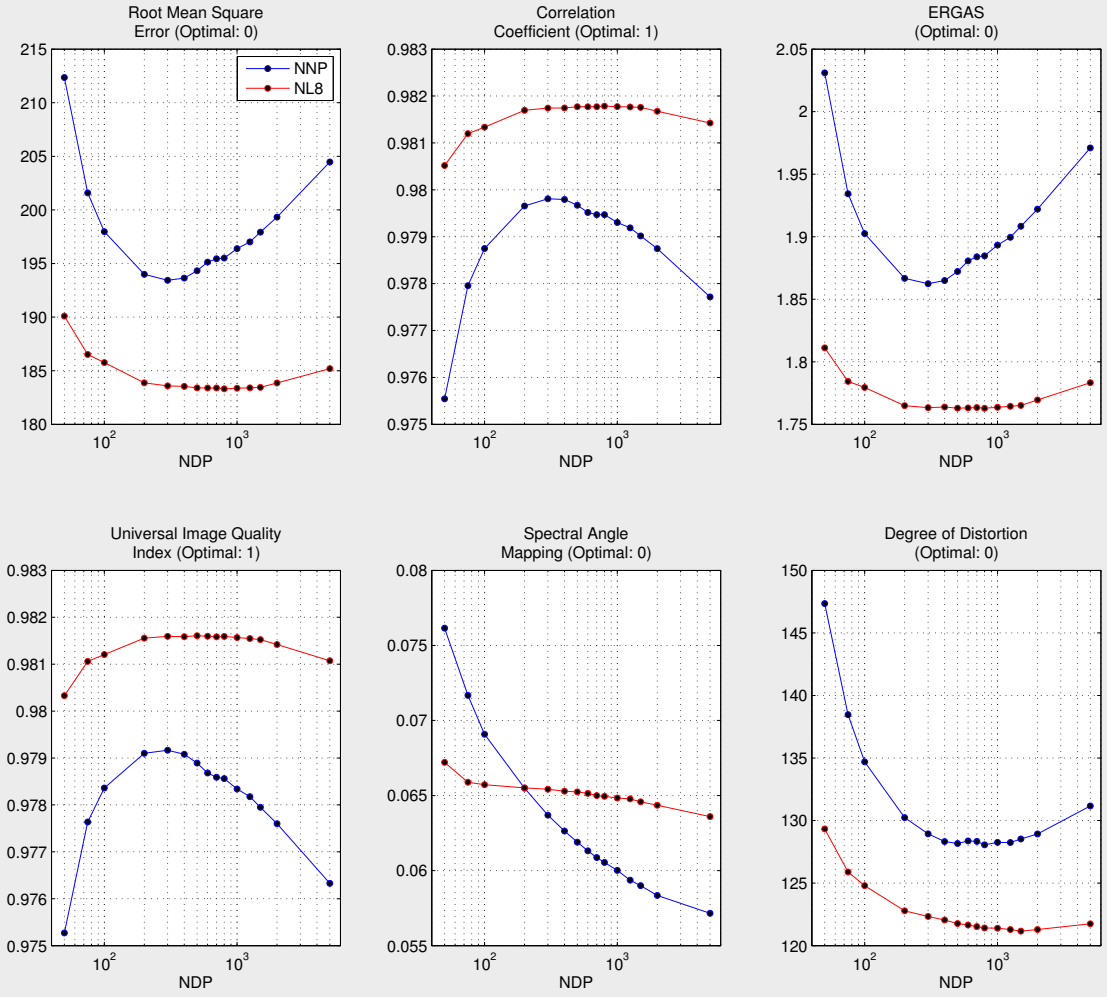
Figure 5.32: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL8 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.10.1 NL8 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.33.

Figure 5.33: The performance of the NL8 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



As is evidenced by Figure 5.33, the NL8 dictionary selection method shows a considerable improvement for all image quality metrics, with the exception of the SAM. It should be noted that this improvement is of sufficient magnitude that even the performance of the smallest tested dictionary, $N_{DP} = 50$, outperforms or matches the optimal benchmark performance at

$N_{DP} \approx 300$. The SAM metric varies less than the benchmark result, resulting in inferior SAM values for $N_{DP} > 200$. Furthermore, the respective peaks and troughs for the NL8 dictionary method are more shallow than the NNP benchmark, indicating that NL8 dictionary selection provides more robust performance over a greater range of N_{DP} values.

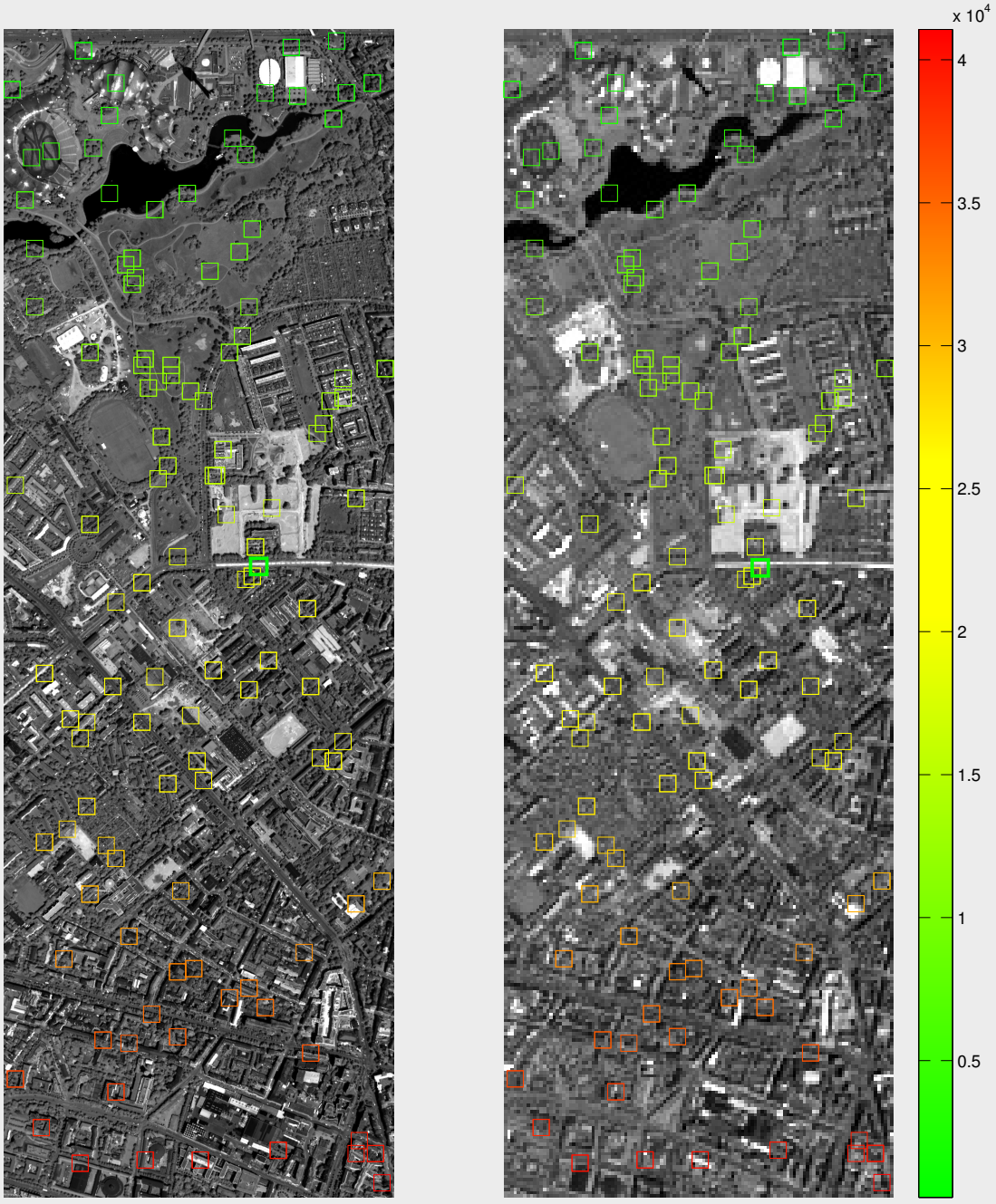
5.11 Non-Local PanHR Random Dictionary Selection (NL9)

In order to provide a control measurement for all assessment metrics, including the benchmark NNP method, random patches were taken to form random coupled dictionaries.

Note however, there is one *crucial* subtly to the NL9 random dictionary selection method. The first element in the dictionary was taken to correspond to the current patch under reconstruction i.e. the *ith* PanHR and PanLR patches, corresponding to the *ith* patch under reconstruction, are respectively included as the first element in the HR and LR dictionaries respectively.

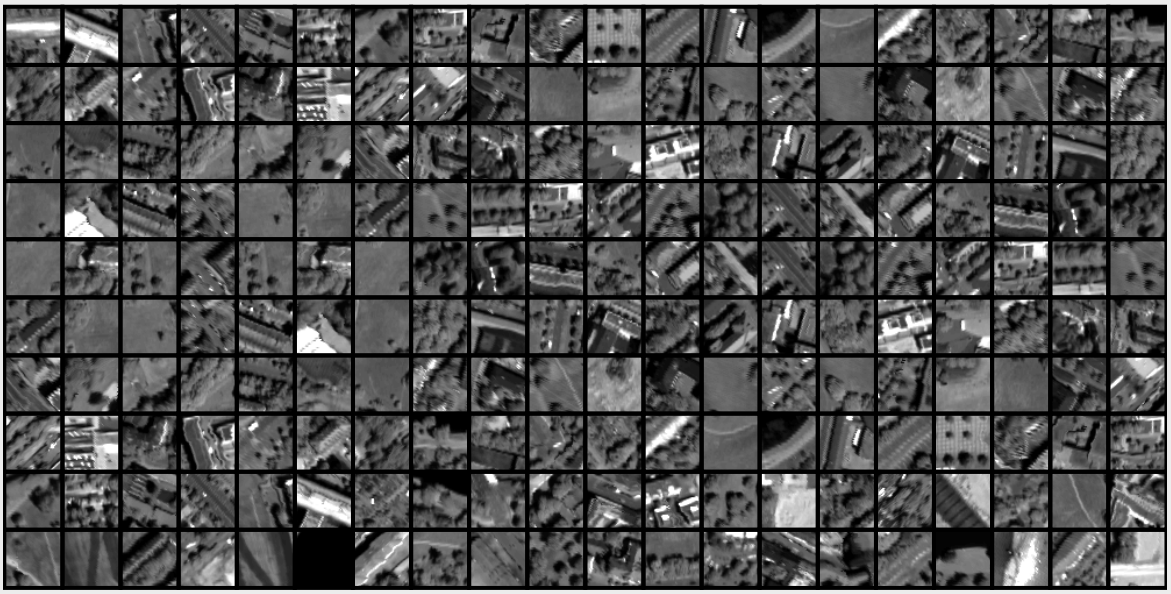
Note that valid random or pseudo random number generation in parallel computing can be problematic, due to difficulties in effectively seeding random functions. For this implementation a random seed is generated by combing the current patch index and current processor time, according to mpi writetime function, into a pseudo-hashing function. While this proved adequate for the purposes of testing, more optimal and statistically valid random implementations should be investigated. The NL9 dictionary location image is displayed in Figure 5.34.

Figure 5.34: The NL9 coupled dictionaries of size $N_{DP} = 200$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents index of the patch, with 0 denoting the first patch in the image.



The complete NL9 dictionary for $N_{DP} = 200$ is displayed in Figure 5.35.

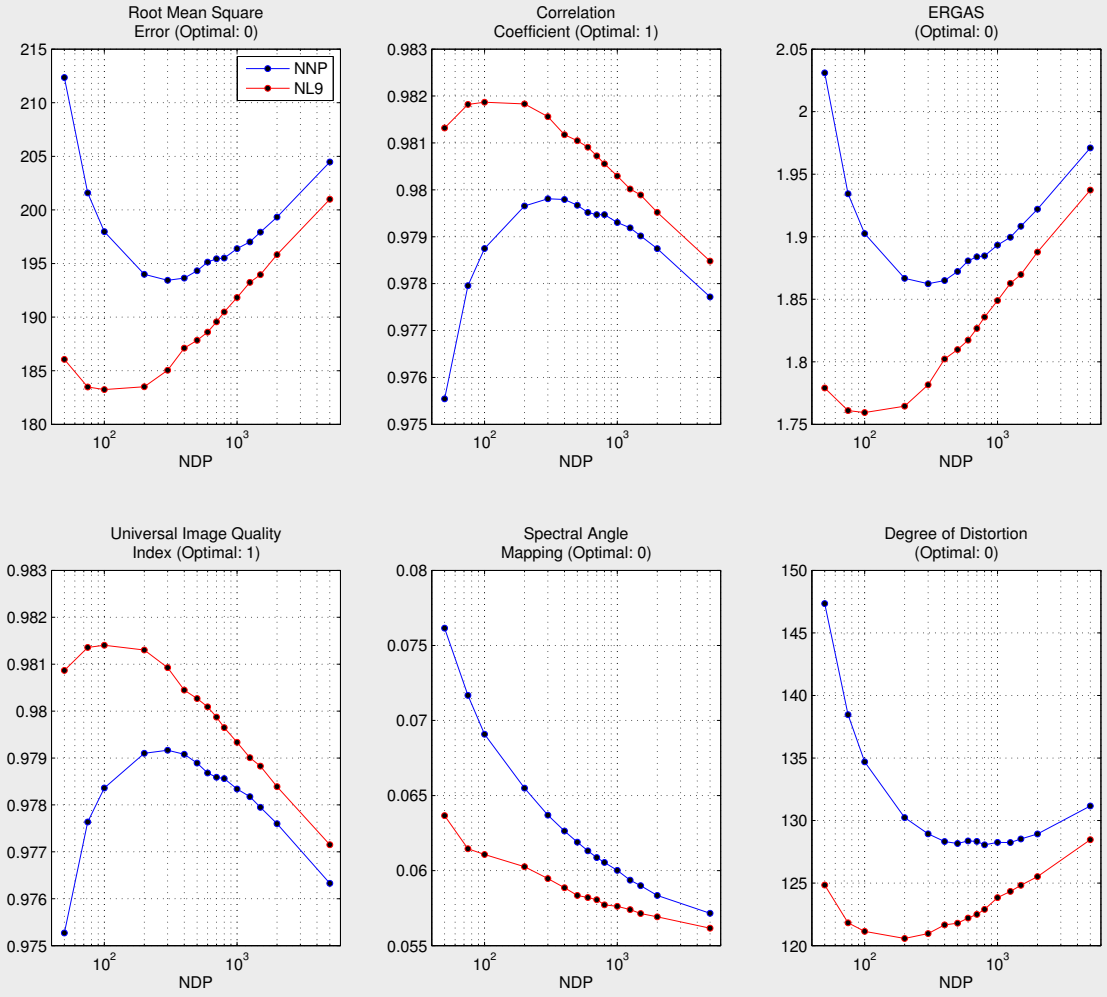
Figure 5.35: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL9 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



5.11.1 NL9 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3, with the same range of N_{DP} values as found in Section 5.2.2. The results are compared to the benchmark NNP dictionary selection in Figure 5.36.

Figure 5.36: The performance of the NL9 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



Considerable improvement on the benchmark NNP selection is observed for all assessed performance metrics using the NL9 random dictionary selection method. This is a highly notable result, given the tradeoff between SAM and spatial based metrics observed in all earlier trials. Peak performance is achieved at $N_{DP} \approx 100$. It should be noted that while

there is an improvement on the SAM metric with respect to the benchmark NNP, NL9 does not match the SAM values for HR similarity based measurements such as NL7, NL6, NL5 and NL3.

5.12 Non-Local PanHR Self-Uncorrelated Dictionary Selection (NL10)

The motivation behind the Non-Local PanHR Self-Uncorrelated Dictionary Selection (NL10) method was an attempt to exploit a coupled dictionary structure containing a basis set of atoms. An ideal basis can represent every vector in its given vectorspace as some linear combination, thus, theoretically, better facilitating reconstruction of any HRMS patch under reconstruction.

Since a basis is simply a set of linearly independent vectors, a method for obtaining an approximate basis set would be to select HR patches on the basis of cumulative anti-correlation to the HR patches already in the HR dictionary, forming a "self uncorrelated" dictionary. As with previous NL8 and NL9 dictionary selection methods, the first atom in the dictionary was taken to correspond to the current patch under reconstruction i.e. the *ith* PanHR and PanLR patches, corresponding to the *ith* patch under reconstruction, are respectively included as the first element in the HR and LR dictionaries respectively.

Given this initial patch in the HR dictionary, a self-uncorrelated dictionary is built up in the following manner:

NL10 Dictionary Selection Algorithm

---

1. The first atom in the HR dictionary, $\mathbf{D}_h^n$, is designated as the *ith* PanHR patch,

corresponding to the reconstruction of the *ith* MSHR patch.

2. For I = 1 to I = N_{DP}

   Note: I = current dictionary size

   a) For J = 1 to J = I

      Loop n through all N_P−I patches which are not currently selected. Assign *cumulative* anti-correlation to the J*th* patch in accordance with:

$$\mu^n_{NL10} += \mu^{\mathrm{J},n}_{NL8} \tag{5.17}$$

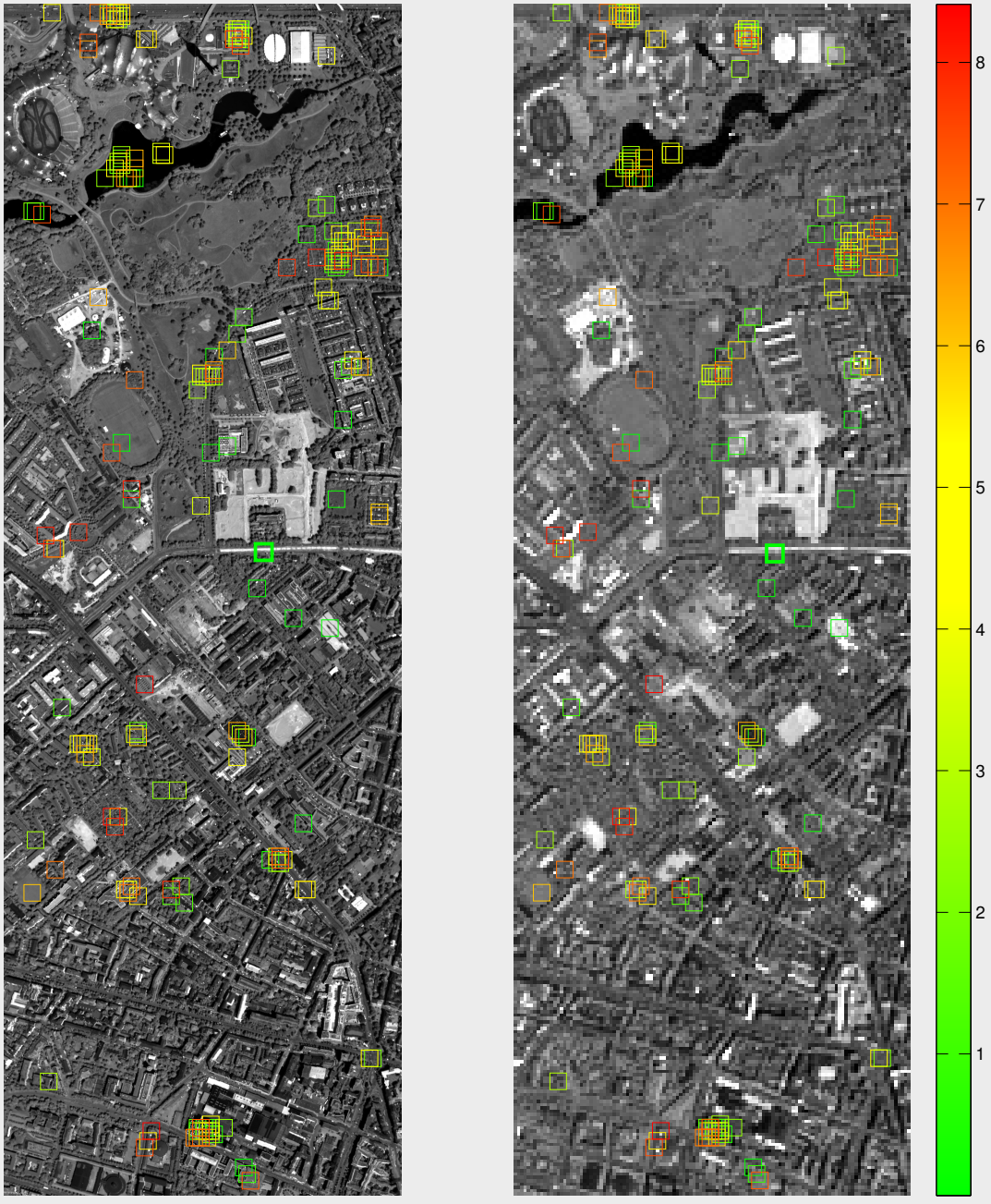
   Add the *nth* PanHR patch to the HR dictionary which is the most *cumulatively* anti-correlated, in accordance with μ_{NL10}, as per Equation (5.17).

3. Patches are selected for the coupled LR dictionary, $\mathbf{D}^n_l$, that directly correspond to the HR dictionary, $\mathbf{D}^n_h$.

The numerical expense of this method is apparent in the abundance of nested loops. To give an example, for a dictionary of size $N_{DP} = 300$, and a total number of patches $N_P = 41296$ there are $\approx 299! \times 40000$ correlation computations performed! Essentially, in searching for a basis set, the NL10 dictionary selection algorithm performs a *very expensive* combinatorial search. This computational expense manifests itself as a limitation on the size of the dictionary that can realistically be tested, even with access to the SuperMUC supercomputing facilities.

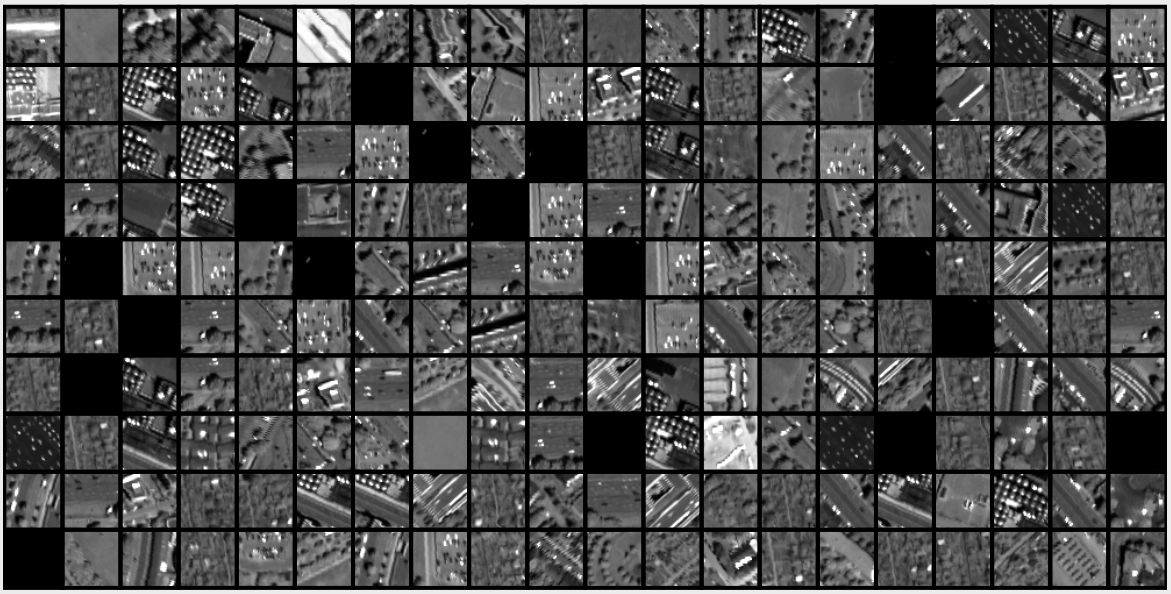
The NL10 dictionary location image is displayed in Figure 5.37.

Figure 5.37: The NL10 coupled dictionaries of size $N_{DP} = 300$ are depicted for an example road patch under reconstruction. The patch locations for the coupled $\mathbf{D}_h^n$ and $\mathbf{D}_l^n$ are arranged on the respective PanHR (left) and PanLR (right) images. The example patch under reconstruction is indicated by the green square, and the colorbar represents the cumulative anti-correlation i.e. μ_{NL10}.



The complete NL10 dictionary for $N_{DP} = 200$ is displayed in Figure 5.38.

Figure 5.38: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NL10 selection method. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right. Note that the first patch in the dictionary corresponds to the current patch under reconstruction.



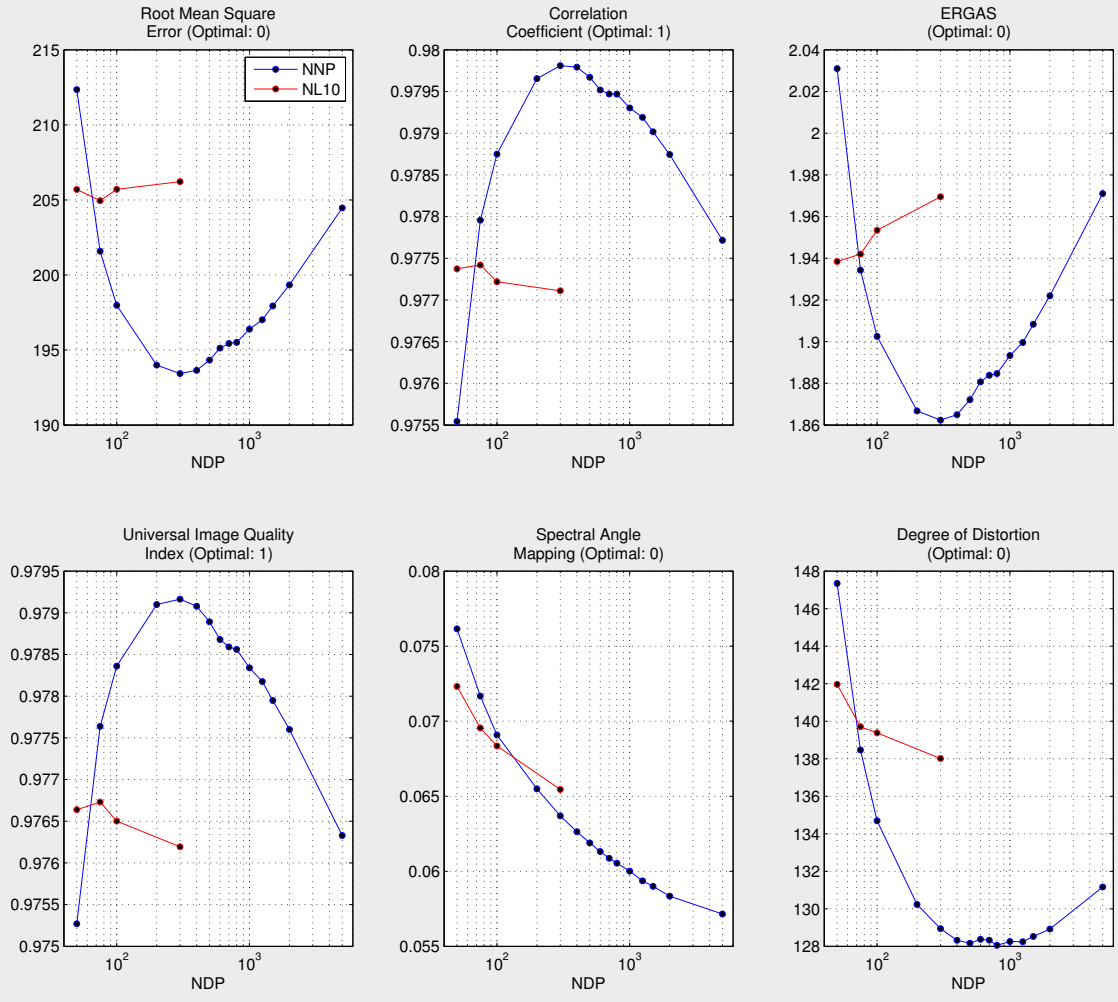
Note the prevalence of point or spot features, which is consistent with the selection of cumulative anti-correlation. This is due to patches characterised by point features that do not *exactly* align with those of patches already in the dictionary having a very low mutual correlation, despite the scene similarity observed.

5.12.1 NL10 Performance

Quantitative performance testing was conducted on the given reconstruction area depicted in Figure 4.3. Due to the aforementioned computational expense only N_{DP} values of 50,

75, 100 and 300 were tested. The results are compared to the benchmark NNP dictionary selection in Figure 5.39.

Figure 5.39: The performance of the NL10 dictionary selection method (red) is assessed for varying N_{DP}, and compared with the benchmark NNP dictionary selection method (blue). Image quality metrics are detailed in Appendix A.



As is depicted in Figure 5.39, mediocre performance was observed for the dictionary sizes tested. In fact, for the $N_{DP} = 300$ trial, performance was considerably worse than the NNP

benchmark, for all metrics excluding the SAM, which was comparable. However given the limited data acquisition, due to a combination of expensive cpu hour resource allocation and time required for computation, there are limits on the conclusions that can be drawn from these trials.

In Chapter 6, the mediocre observed performance is argued to be due to the presence of a speckle/spot artifact, a byproduct of the favourably selected point of spot patches, as mentioned above. Given the excessive and expensive computational time required and the mediocre performance observed, under these particular J-SparseFI reconstruction settings with this particular dataset, further implementation or testing of the NL10 self-uncorrelated selection method is not recommended.

Chapter 6

Dictionary Selection Evaluation

This section provides evaluation and critical analysis regarding the performance of the dictionary selection methods implemented in Chapter 5. The trade-off between spectral and spatial fidelity, with regards to dictionary selection techniques, is discussed and evaluated in terms of the respective sparse reconstruction coefficients. A computational efficiency estimate is provided for each of the assessed dictionary selection methods. Recommendations are made for dictionary selection techniques based on spatial performance, robustness and spectral fidelity.

6.1 Performance Evaluation

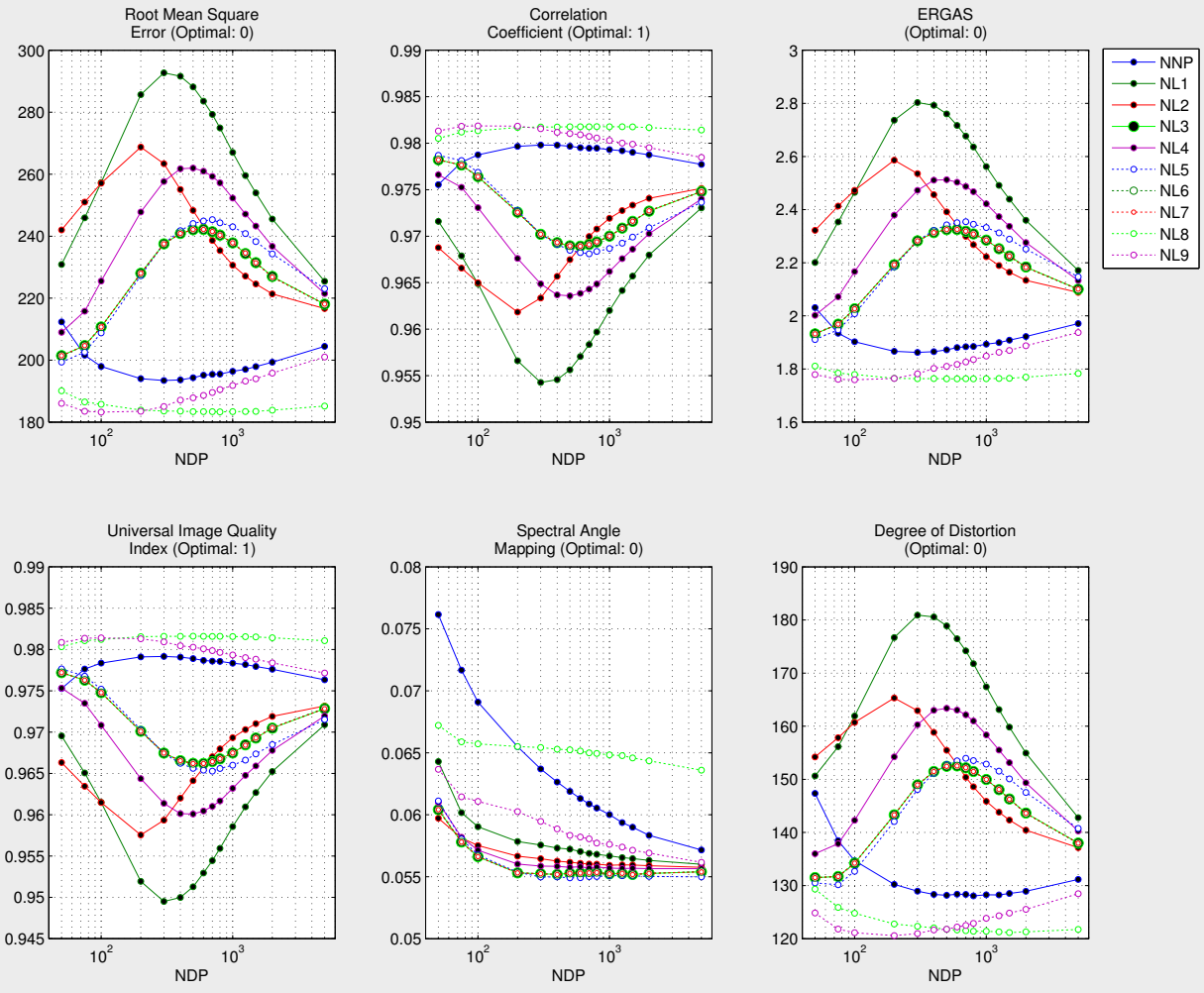
Image quality assessment graphs depicting the relative performance of each of the investigated dictionary selection methods are depicted in Figure 6.1. The dictionary selection methods fall into two broad categories:

1. Selection methods that select patches based on similarity to the current patch undergoing reconstruction. These include NL1, NL2, NL3, NL4, NL5, NL6 and NL7.

2. Selection methods that select patches that are probabilistically dissimilar to the patch undergoing reconstruction. These include NNP, NL8, NL9 and NL10. While the NNP benchmark method selects patches in the surrounding neighborhood, with probable scene similarity, geometrical correlation with the patch under reconstruction is unlikely,

given the possible periodic effects of selecting patches in rings located at ever increasing distances from the current patch.

Figure 6.1: Dictionary Selection Performance is evaluated for the standard image quality assessment metrics, as detailed in Appendix A.



Note that the NL10 dictionary selection method was not included, due to insufficient data collection related to the excessive computational times involved.

Figure 6.1 displays a number of highly significant points worthy of detail:

- The random dictionary selection method, NL9, achieves peak performance for all image assessment metrics, excluding SAM, at a dictionary size of $N_{DP} \approx 100$.
- The random dictionary selection method, NL9, is the only tested method that manages an improvement over the benchmark in the spacial based metrics and SAM.
- The most robust performance in all metrics, excluding SAM, is provided by the uncorrelated dictionary, NL8. this is illustrated by the shallow curve, with minimal variation for different dictionary sizes.
- Similarity based selection methods perform poorly in spatial metrics, yielding a characteristic peak or trough corresponding to the respective minimum performances in spatially weighted image quality metrics.
- The NL3, NL6 and NL7 curves overlap exactly, due to their mathematical equivalence in this particular problem, as previously mentioned in Chapter 5.
- The peak SAM performance is achieved for the NL3, NL6 and NL7 HR similarity based selection methods. this indicates that similarity based metrics provide reconstructions with the highest spectral fidelity, by a considerable margin compared to the benchmark NNP.
- The peak SAM performance achieved by the NL3, NL6 and NL7 hr is robust with respect to dictionary size, with minimal variation for $N_{DP} \geq 200$.
- The peak SAM performance achieved by NL3, NL6 and NL7 appears to correspond to the SAM values that the other metrics converge on with increasing N_{DP}.

It is also important to consider the reconstructed images to evaluate reconstruction quality and assess the presence and severity of artifacts, blurring, tone and other perceptual qualities.

The reconstructed HRMS images for the most competative dictionary selection methods, in addition to the LRMS measurement image and HRMS reference image, are depicted in Figure 6.2. All other reconstructions can be found in Appendix B.1.

Figure 6.2: Here the LRMS measurement image that is sharpened is depicted in a), along with the HRMS reference image b). The other depictions are; benchmark NNP dictionary selection reconstruction c), NL9 Random dictionary selection d), NL8 Anti-correlation dictionary selection e), and NL7 patch angle mapping dictionary selection f).

The reconstruction quality of the benchmark NNP dictionary selection method is apparent in Figure 6.2 c). Note however the presence of spot artifacts in the vicinity of vegetation

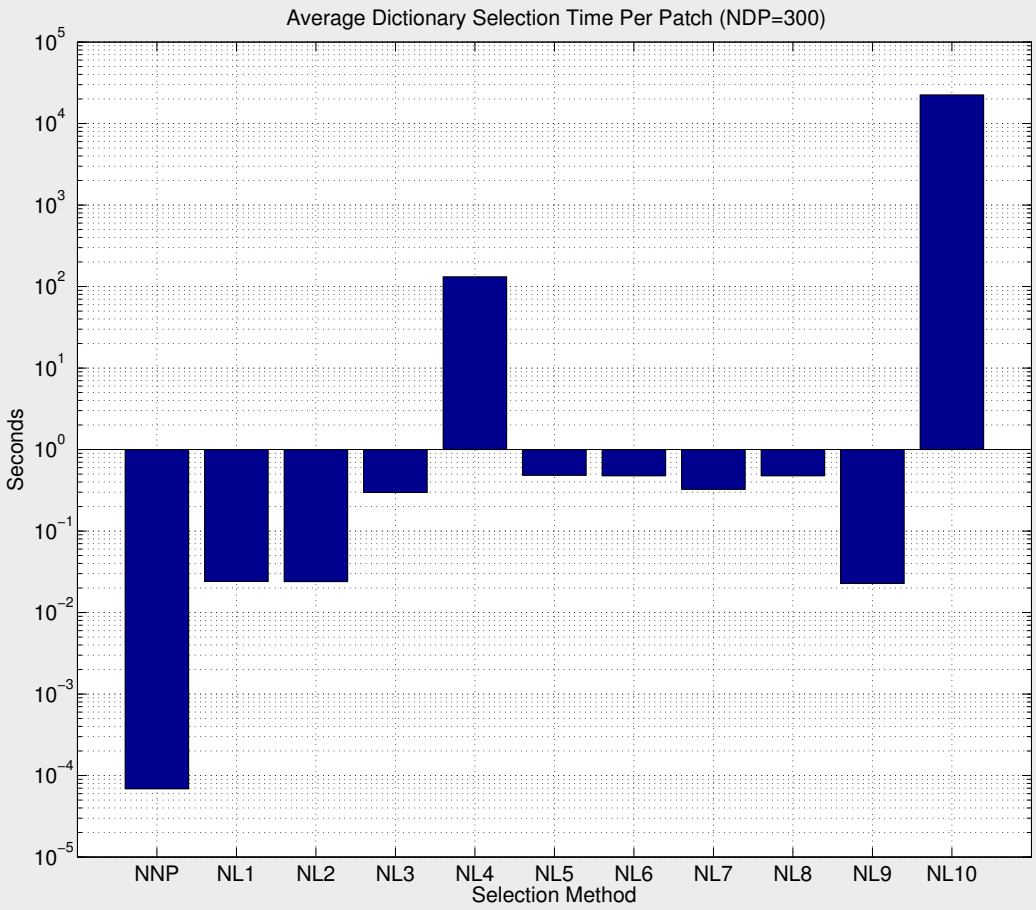
textures. This is explained by the nearest neighbor method selecting a greater proportion of vegetation patches in the case that the patch undergoing reconstruction is in the locality of vegetation structures. Given that the vegetation, in particular trees, maintains a circular structure when viewed from above, this circular structure manifests itself as a spot/speckle artifact in the final reconstruction. This artifact is eliminated by choosing a random dictionary with a low probability of including a high proportion of vegetation patches (given the current dataset), as demonstrated in Figure 6.2 d). The elimination of this artifact using the random NL9 dictionary selection method likely accounts for the improvement seen in all performance metrics, when compared with the benchmark NNP.

Also note the sharpness and lighter tone of the NL8 reconstruction in Figure 6.2 e). In fact, in comparison to the HRMS reference image in Figure 6.2 b), the vegetation appears visually sharper, beyond the natural sharpness observed in the reference image. The difference in visual tone and relative sharpness is consistent with the reduced spectral performance and robust spatial performance measured for NL8.

Immediately apparent in Figure 6.2 e), the NL7 patch angle similarity based measurement, is the relative blurriness of the reconstruction, evidence of HR detail loss. Also notable is the tone of the image, particularly with respect to grassy areas and vegetation, which better matches the HRMS reference image in Figure 6.2, in comparison with the other reconstructions. This result is also consistent with the observed metrics in Figure 6.1.

Appropriate analysis of the dictionary selection methods also demands consideration of their relative computational times. Figure 6.3 displays the average dictionary selection computational time per patch, for dictionaries of size $N_{DP} = 300$. Further note that the average time for NL10 dictionary selection is also included, to demonstrate the excessive computational times required for selection, $\approx 10^4$seconds.

Figure 6.3: Here the average dictionary selection time per patch is depicted for the assessed metrics. Note that the time values are in seconds, and plotted on a log scale.



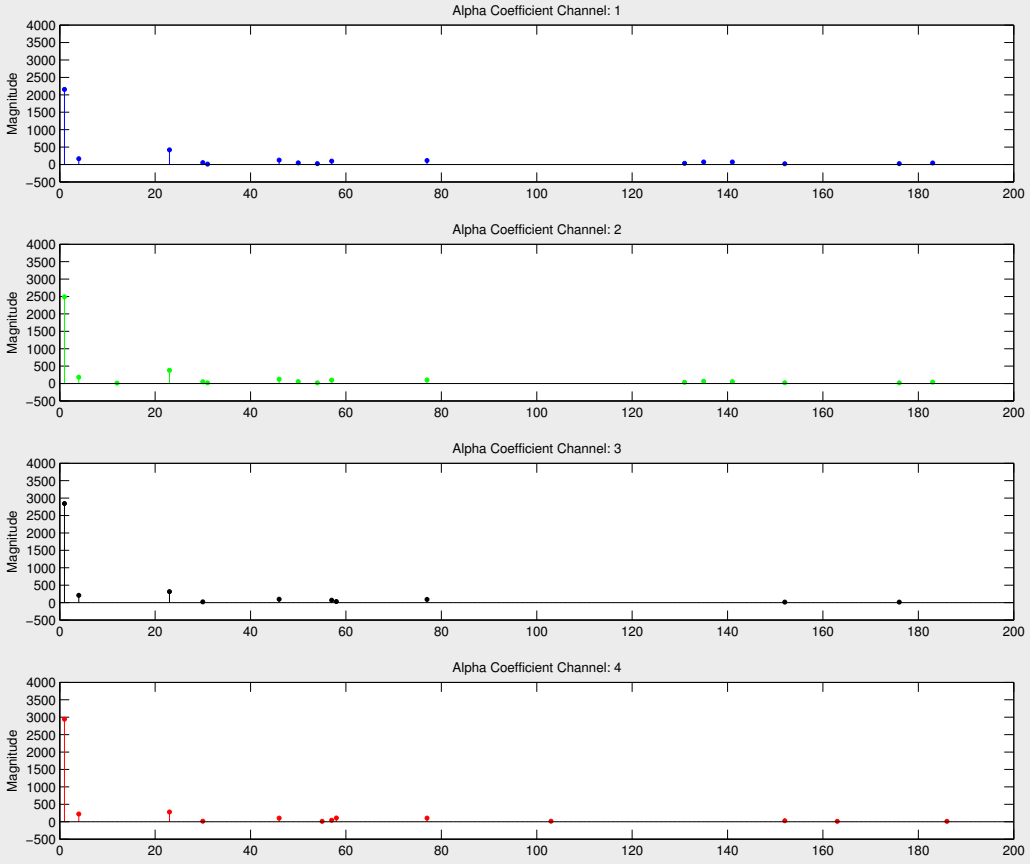
It is important to note that NNP dictionary selection is at least two orders of magnitude faster than all other dictionary selection methods implemented. Furthermore there is an order of magnitude increase in time associated with the HR comparison methods, NL3, NL5, NL6 and NL8, when compared to the LR norm comparison methods, NL1 and NL2. Also note the modest time saving associated with both the patch angle mapping NL7 selection and HR norm NL3 selection, which both provide identical performance to the positive correlation based NL6 method. Also note that the random dictionary selection is relatively fast,

with times comparable to the LR norm comparison based methods.

The fact that all implemented NL dictionary selection methods have not been optimised for speed or memory performance must also be taken into account. Figure 6.3 is instead intended to provide relative comparison.

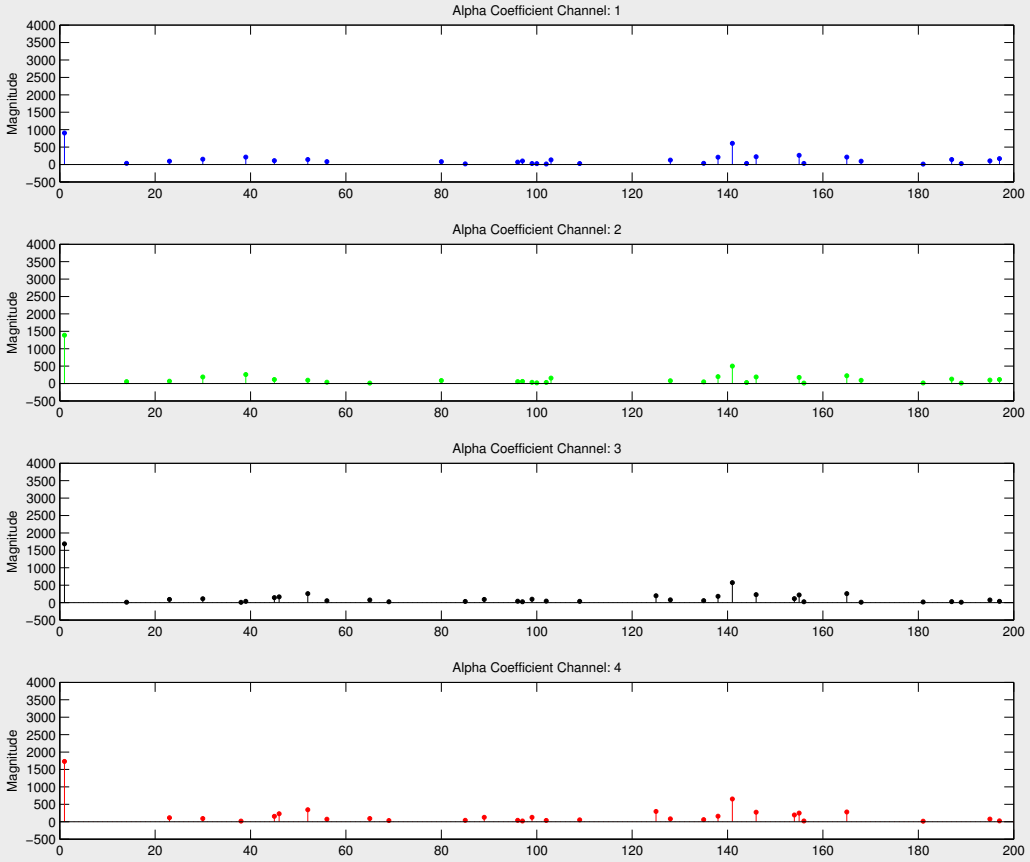
In order to provide insight into the differences in performance for the different dictionary selection categories the sparse reconstruction coefficient values for the reconstructed channels are graphed. Figure 6.4 depicts the alpha coefficients for the NNP benchmark selection method.

Figure 6.4: The sparse reconstruction coefficients, α, are graphed for the reconstructed channels 2-5, for the NNP benchmark. The colours red, green and blue correspond to the channels respective approximate visual wavelengths



Note the highest magnitude coefficient corresponds to the first atom in the dictionary. Since the first dictionary atom corresponds to the patch that is undergoing reconstruction, this is an expected result, as the information required for reconstruction is predominantly present in the respective panchromatic region. The NNP sparse coefficient magnitude distribution is in strong contrast to the sparse coefficient distribution of the similarity based dictionary selection methods. Figure 6.5 displays the sparse reconstruction coefficient distribution for an exemplar similarity based selection method, the patch angle mapping NL7.

Figure 6.5: The sparse reconstruction coefficients, α, are graphed for the reconstructed channels 2-5, for the NL7 selection method. The colours red, green and blue correspond to the channels respective approximate visual wavelengths

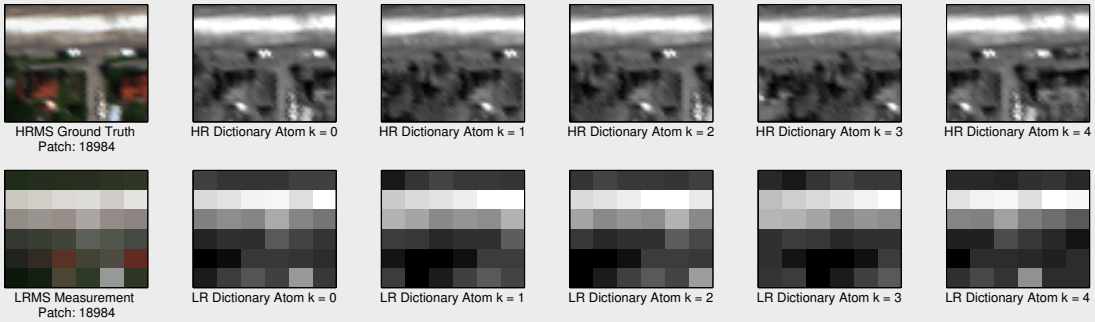


Note that the other similarity based selection alpha coefficients, which are detailed in Appendix B.2, also display more equally distributed magnitudes.

Immediately noticeable is a reduction in the peak of the weighting of the first dictionary atom in the reconstruction, and a consequent increase in the weighting of other dictionary atoms. This corresponds to the alpha coefficients appearing less sparse, in comparison to the NNP benchmark. While, technically speaking, the sparsity of the alpha coefficients is

regulated by the λ parameter, which weights the $L_{2,1}$ penalty term in the regularised sparse optimisation (see Equation (3.17)), apparent sparsity is observed to decrease, due to the relative increase in weighting of dictionary elements. The reduction in apparent sparsity is a consequence of more patches being considered as crucial crucial to the reconstruction by the solver in the LR system. This is illustrated by the considering in detail the LR dictionary atoms of a similarity based selection method.

Figure 6.6: Here the first five patches for the HR dictionary, $\mathbf{D}_h$, and LR dictionary, $\mathbf{D}_l$, for the NL7 selection method are depicted.



Given the visual similarity between the LR atoms in similarity based coupled dictionaries, it is imaginable that the solver selects relevant patches in the dictionary with some difficulty, and accordingly weights dictionary atoms with a more equitable distribution.

It is argued that the increased weighting of additional patches, resulting from a greater proportion of similar patches being included in the coupled dictionaries, allows for greater pixel based refinement in the LR system. This pixel based refinement is necessary for making small corrections and adjustments to the spectral angle mapping for different pixels in the image. This provides a possible explanation for the outstanding SAM results and spectral fidelity of the similarity based selection methods.

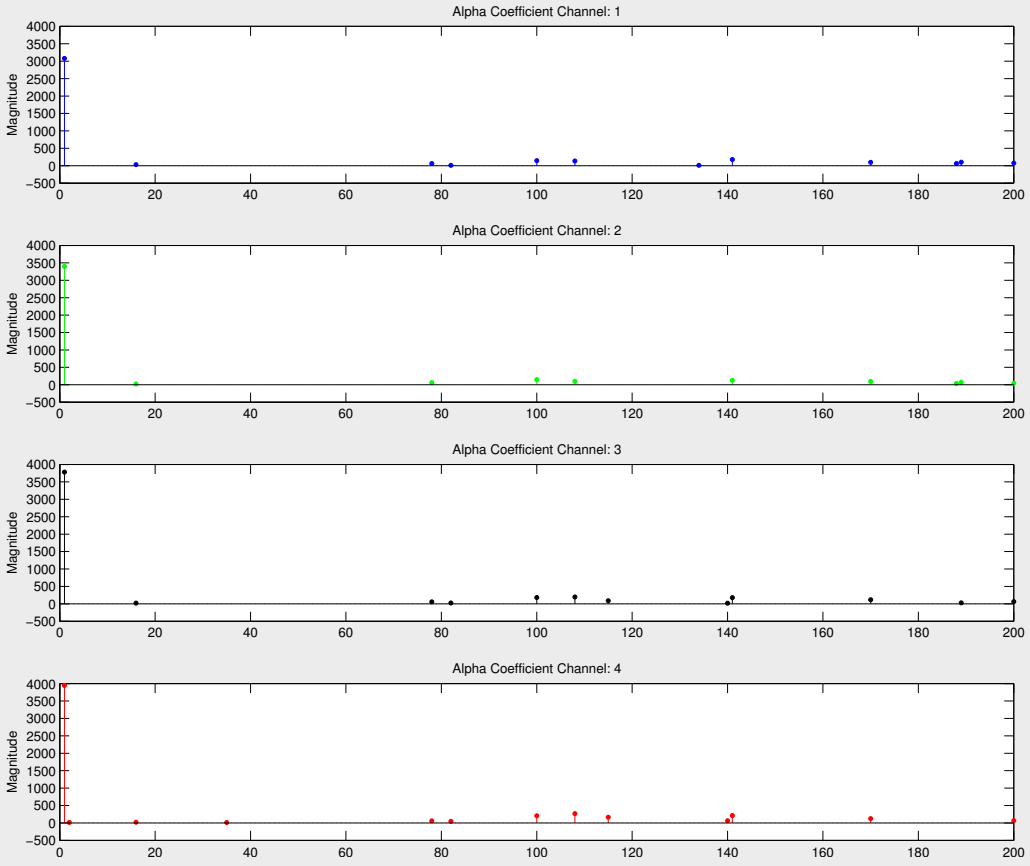
However this same mechanism is likely also responsible for the degradation in spatially

weighted image quality metrics, characteristic of the similarity based selection methods. Since a greater proportion of patches are weighted with higher significance to the reconstruction in the LR system, this translates to correspondingly high weighting int the HR system. Referring to Figure 6.6, it is however evident that geometrical variation, that is only prominent in the HR dictionary atoms, subsequently attains a greater weighting in the reconstruction. This results in the accumulation of geometrical errors, providing explanation for the characteristic poor spatial performance of the similarity based metrics. Explained in another way, geometrical errors in the HR system receive a higher weighting, due to the apparent similarity in the LR system, where the sparse reconstruction coefficient is calculated. This phenomenon also clarifies why the LR norm similarity selection metrics, NL1 and NL2 are outperformed by the HR norm similarity selection metrics, NL3, NL5, NL6 and NL7. Selecting similar patches on the basis of the HR system better controls the geometric variation in the HR system, reducing the accumulation of errors.

This line of reasoning is supported when considering the alpha values of an exemplar probabilistically dissimilar dictionary, the uncorrelated NL8.
Figure 6.7 displays a considerable increase in the magnitude of the weighting of the first dictionary element, for all channels. Note that a critical feature of the probabilistically dissimilar dictionary selection methods, NNP, NL8, NL9 and NL10 was the inclusion of the PanLR and PanHR patches corresponding to the current patch under reconstruction. Without this inclusion, exceedingly poor performance is observed. This is expected, given the geometrical information for sharpening is predominantly provided by the corresponding location in the Pan image.

Thus given, it is expected that the initial dictionary element is weighted with increased significance in the NL8 uncorrelated dictionary. Given that there is considerably less likelihood for patches to be considered significant in the LR system, this translates to increased weighting of only the patch corresponding to the current reconstruction in the HR system. This

Figure 6.7: The sparse reconstruction coefficients, α, are graphed for the reconstructed channels 2-5, for the NL8 selection method. The colours red, green and blue correspond to the channels respective approximate visual wavelengths



allows for greater preservation of the HR geometric information contained in the PanHR image. However, as reasoned above, this corresponds to a reduced capacity for pixel based refinement in spectral fidelity, reflected in the comparatively poor SAM results yielded by the NL8 method, and the probabilistically uncorrelated dictionary selection methods in general.

6.2 Conclusion and Recommendations

There is a complex tradeoff between spectral fidelity and spatial accuracy when considering dictionary selection methods for J-SparseFI. Given the assessment of sparse reconstruction coefficients, this relationship is based around a combination of the LR dictionary patches passed to the LR sparse optimization solver and the spectral and geometrical validity of this weighting in the HR system. It should be noted that in all cases, inclusion of the HR and LR Pan patches corresponding to the patch undergoing reconstruction was *mandatory* for achieving competitive reconstruction performance.

In general, similarity based selection methods demonstrate a characteristic curve in spatial based performance metrics, with considerably worse performance than the benchmark NNP. Similarity based methods, particularly HR similarity based selection, provide outstanding SAM performance, indicating a high degree of spectral information preservation. It is argued that this spectral fidelity is a result of the a more even distribution of sparse coefficient weightings, allowing for more pixel based correction, improving the average spectral angle mapping. This more even coefficient magnitude distribution is inferred to be due to LR dictionary atom similarity, resulting in the solver assigning a higher weighting to more dictionary atoms for reconstruction. Poor spatial performance for similarity based selection methods is attributed to the accumulation of geometrical errors in the HR system, due to the aforementioned increase in relative dictionary atom weightings in the sparse coefficients.

Probabilistically dissimilar dictionaries, by contrast, demonstrate consistently better spatial performance. Following the reasoning above, dissimilar dictionaries are characterised by considerably higher sparse coefficient weighting of the first atom in the coupled dictionaries, corresponding to the current reconstruction area. This result is intuitive, given that the LR dictionary probabilistically contains greater diversity, allowing for greater solver patch discrimination and consequent concentration of sparse coefficient weighting around the first

dictionary atom, corresponding to the current reconstruction patch location. Given the high weighting of the first dictionary atom, a high degree of spatial information encoded in the PanHR image is preserved in the reconstruction, accounting for the better spatial performance. However, the reduction in weighting for additional dictionary atoms reduces the potential for pixel based corrections to spectral fidelity, evident in the considerably poorer SAM performance of probabilistically dissimilar dictionaries.

Worthy of note, however, is the mediocre performance of the NL10 self uncorrelated dictionary, despite the high likelihood of mutual orthogonality of the dictionary atoms. This lack of performance is inferred to be due to the prevalence of patches with point features, which produce low self correlation, but introduce speckle artifacts and fail to capitalize on possible geometric correlations and information contained in the PanHR image. Furthermore, due to the effectively combinatorial search required, NL10 has exceedingly low computational efficiency, with computational times exceeding 10^4 seconds for dictionaries of size $N_{DP} = 300$, *well in excess of the time required for the resolution of the sparse optimisation problem for* $\lambda = 1$.

In light of this investigation a number of dictionary selection recommendations are presented. There are evident tradeoffs, particularly regarding spectral and spatial fidelity, that must be accounted for. Thus, user application based recommendations are made for the following general performance categories:

Spatial Performance: Peak performance for all assessed image quality metrics, excluding the SAM, was attained by using the NL9 dictionary selection method, by which the initial dictionary atom in the HR and LR dictionaries corresponded to the current patch under reconstruction and the other atoms were randomly selected from the total set of patches. NL9 also demonstrated improvement over the NNP benchmark with regards to the SAM metric. Furthermore, NL9 was one of the most computationally efficient dictionary selection

methods tested, though it was not as fast as the NNP benchmark.

Robustness: With regards to the dictionary size, N_{DP}, the most robust performance attained for all assessed image quality metrics was recorded for the NL8 anti-correlated dictionary selection method. This method selects the initial dictionary atom in the HR and LR coupled dictionaries to correspond to the current patch under reconstruction, and selects the remaining atoms on the basis of HR anti-correlation to the PanHR patch corresponding to the current patch under reconstruction. Relatively flat performance curves are observed for the NL8 anti-correlated dictionary selection method, indicating robust performance over a wide range of dictionary sizes, N_{DP}. Furthermore, NL8 provided considerable improvement over the benchmark NNP for all image assessment metrics excluding SAM, far all N_{DP}. Furthermore, N8 provides performance enhancement without being subject to the random variation in performance, which is an intrinsic characteristic of the NL9 random dictionary selection method.

Spectral Fidelity: While the highest spectral fidelity, in terms of the SAM performance metric, was jointly attained by NL3, NL6 and NL7 HR similarity based selection methods, due to aforementioned mathematical equivalence in this particular problem, the NL3 and NL7 methods performed better in computational efficiency. Though the computational efficiency of NL3 and NL7 was comparable, the advantage of the Patch Angle Mapping (PAM) method is that it is independent of patch intensity, similar to a correlation based selection method, and thus only geometrically selective. While, for this investigation this intensity independence was not significant, due to patch normalisation, this method is of interest to broader compressive sensing based approaches, which may not necessarily work with normalised, zero mean images. Therefore, the HR similarity based NL7 PAM dictionary selection is recommended for optimal spectral performance.

Recommendations for continued investigation are as follows:

- Characterisation of the variation between trials for the NL9 random dictionary selection method. Included should be an assessment on the statistical validity of the pseudo-hash function used to seed the random number generator.

- Speed and memory optimisation for all implemented dictionary selection techniques, with emphasis on NL9, NL8 and NL7.

- Additional investigation into the impacts of dictionary training for hyperspectral-multispectral image sharpening should be conducted, with particular emphasis on similarity based selection, due to the *substantial* improvement in spectral fidelity observed in the pan-sharpening problem.

Part II

Dictionary Training

Chapter 7

K-SVD for J-SparseFI Dictionary Training

The K-SVD algorithm trains a dictionary from a set of signals, by updating atoms to match the most significant source of error in the SVD decomposition of the reconstruction without the respective atom. Here an algorithm is proposed which trains a dictionary to best represent the jointly sparse MS image, reconstructed using J-SparseFI. Alterations to the algorithm are also introduced, specifically a reposed problem in which information from the LR system is concatenated to the HR system, and a reduced system, potentially improving computational efficiency.

7.1 Introduction to K-SVD

K-SVD is an iterative dictionary training method that alternates between optimisation of the sparsity coefficient for the current dictionary and dictionary atom update based on improving the fit to the data [14]. By combining the update of the dictionary with the sparse coefficient update, convergence is typically accelerated. The K-SVD algorithm is flexible and can be tailored to most pursuit methods [14], allowing for diverse applications in signals analysis, including pan-sharpening [16].

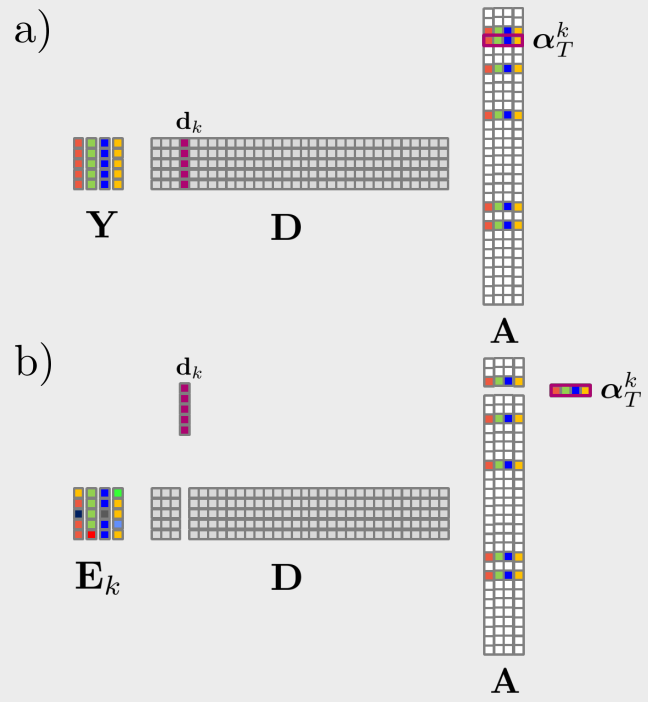
The general operating principle behind the K-SVD model can be described by considering

the training of a dictionary, $\mathbf{D}$, from a set of training signals, $\mathbf{Y}$, with a sparse reconstruction coefficient, $\mathbf{A}$. The optimization of the dictionary and the sparse reconstruction coefficient are alternatively updated. Following the initialisation of the sparse reconstruction coefficient, the update of the k^{th} dictionary element is posed as the following penalty term:

$$\begin{aligned} \|\mathbf{Y} - \mathbf{DA}\|_F^2 &= \left\| \mathbf{Y} - \sum_{j=1}^{K} \mathbf{d}_j \boldsymbol{\alpha}_T^k \right\|_F^2 \\ &= \left\| \left(\mathbf{Y} - \sum_{j \neq k}^{K} \mathbf{d}_j \boldsymbol{\alpha}_T^j \right) - \mathbf{d}_k \boldsymbol{\alpha}_T^k \right\|_F^2 \\ &= \left\| \mathbf{E}_k - \mathbf{d}_k \boldsymbol{\alpha}_T^k \right\|_F^2 \end{aligned} \tag{7.1}$$

Where $\boldsymbol{\alpha}_T^j$ is the j^{th} *row* of $\mathbf{A}$. Hence, $\mathbf{E}_k$ represents the error in reconstructing the training set with all dictionary atoms *except* the k^{th} atom. This is illustrated in Figure 7.1:

Figure 7.1: a) depicts set of training signals, $\mathbf{Y}$, that are used to train a dictionary, $\mathbf{D}$, which sparsely represents the training signals according to a sparse coefficient matrix $\mathbf{A}$. In b), an error term, $\mathbf{E}_k$, is constructed, corresponding to the reconstruction of the training signals excluding the *kth* dictionary element, and corresponding sparse coefficients.



Likewise, $\mathbf{d}_k\boldsymbol{\alpha}_T^k$ represents the contribution of the k^{th} dictionary atom to the final reconstruction. Given that $\mathbf{A}$ is sparse, it follows that the rows, $\boldsymbol{\alpha}_T^k$, are likely also sparse, except in the special case of joint sparsity. Thus, generally, the system is reduced to:

$$\left\|\mathbf{E}_k^R - \mathbf{d}_k\boldsymbol{\alpha}_R^k\right\|_F^2 \tag{7.2}$$

Where R denotes the reduced error terms and sparse coefficients that do not contain zero entries. It follows that this reduced penalty term can be minimised in the case that $\mathbf{d}_k$ and

$\boldsymbol{\alpha}_R^k$ are selected to cancel the most significant contribution of the error. This can be done by considering the Singular Value Decomposition of the $\mathbf{E}_k^R$ term.

Definition 7.1. *For any given matrix* $\mathbf{E} \in \mathbb{R}^{m \times n}$ *there exists a Singular Value Decomposition (SVD):*

$$\mathbf{E} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \tag{7.3}$$

Such that:

- $\mathbf{U}$ *is an* $m \times n$ *matrix with orthogonal columns. The columns of* $\mathbf{U}$ *are known as the Left Singular Vectors of* $\mathbf{E}$*.*
- $\boldsymbol{\Delta}$ *is an* $n \times n$ *diagonal matrix with non negative entries. The diagonal values of* $\boldsymbol{\Delta}$ *are called the singular values of* $\mathbf{E}$*. The magnitude of the singular values decrease down the diagonal i.e. the highest magnitude singular value is* $\delta_{1,1}$*.*
- $\mathbf{V}^T$ *is an* $n \times n$ *orthogonal matrix. The columns of* $\mathbf{V}$ *are known as the Right Singular Vectors of* $\mathbf{E}$*.*

Taking the SVD decomposition of the reduced error:

$$\mathbf{E}_k^R = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \tag{7.4}$$

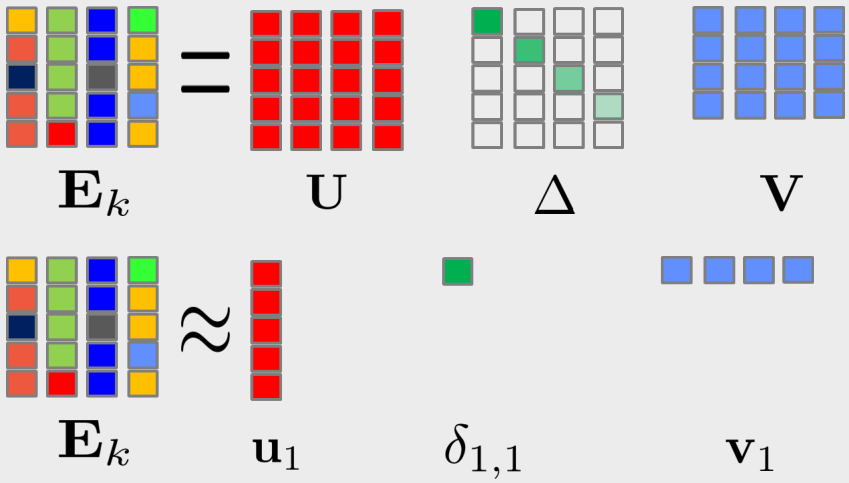
Given that the singular value, $\delta_{1,1}$, represents the *most significant* contributions to the decomposition, the *most significant* sources of error can be negated in a straightforward manner by setting the updated dictionary, $\hat{\mathbf{d}}_k$, as the first column of $\mathbf{U}$, and the sparse reconstruction coefficient, $\hat{\boldsymbol{\alpha}}_R^k$, as the first column of $\mathbf{V}$ multiplied by $\delta_{1,1}$:

$$\hat{\mathbf{d}}_k = \mathbf{u}_1 \tag{7.5}$$

$$\hat{\boldsymbol{\alpha}}_R^k = \delta_{1,1}\mathbf{v}_1^T \tag{7.6}$$

This concept is depicted in **??**:

Figure 7.2: Approximation of the error term, $\mathbf{E_k}$ using the singular value decomposition.



The dictionary and reconstruction coefficients are iteratively reconstructed until the dictionary sufficiently represents the training data set, through convergence or within some predefined error margin. An adaptation to the K-SVD algorithm, for J-SparseFI integration, is proposed in Section 7.2.

7.2 K-SVD Post Training for J-SparseFI Algorithm

K-SVD dictionary training for the J-SparseFI algorithm is posed as follows, with the notation developed in Section 3.2 and Section 3.3:

J-SparseFI with K-SVD Dictionary Training Algorithm

Set Main Loop

Initialise J=1.

1. **Sparse Reconstruction Coefficient**

   The sparse coefficient, $\mathbf{A}^n$, is obtained patch-wise in accordance with the J-SparseFI algorithm in Section 3.3, through the solution of Equation (3.17):

$$\hat{\mathbf{A}}^n = \arg\min_{\mathbf{A}^n} \left\{ \lambda' \|\mathbf{A}^n\|_{2,1} + \frac{1}{2} \|\mathbf{D}_l^n \mathbf{A}^n - \mathbf{Y}^n\|_F \right\} \tag{3.17}$$

2. **HRMS Image reconstruction**

   The HRMS image patch, $\mathbf{Z}^n$, is reconstructed using the sparse reconstruction coefficient, $\mathbf{A}^n$, obtained in step 1:

$$\mathbf{Z}^n = \mathbf{D}_h^n \hat{\mathbf{A}}^n \tag{3.18}$$

3. **K-SVD Dictionary Training**

   a) Applying the sparse reconstruction coefficient to the *high resolution system*, given the standard J-SparseFI assumption of equal coefficients for the HR and LR system, the following term is minimised, as discussed in **??**:

$$\begin{aligned} \|\mathbf{Z}^n - \mathbf{D}_h^n \mathbf{A}^n\|_F^2 &= \left\| \mathbf{Z}^n - \sum_{j=1}^{N_{DP}} \mathbf{d}_j^n \boldsymbol{\alpha}_T^{n,j} \right\|_F^2 \\ &= \left\| \left(\mathbf{Z}^n - \sum_{j \neq k}^{N_{DP}} \mathbf{d}_j^n \boldsymbol{\alpha}_T^{n,j} \right) - \mathbf{d}_k^n \boldsymbol{\alpha}_T^{n,k} \right\|_F^2 \\ &= \left\| \mathbf{E}_k^n - \mathbf{d}_k^n \boldsymbol{\alpha}_T^{n,k} \right\|_F^2 \end{aligned} \tag{7.7}$$

   Where $\boldsymbol{\alpha}_T^{n,j}$ is the *jth row* of $\mathbf{A}^n$. Hence, $\mathbf{E}_k^n$ represents the error in reconstructing the *nth* HRMS image with all dictionary atoms *except* the *kth* atom. Likewise, $\mathbf{d}_k^n \boldsymbol{\alpha}_T^{n,k}$ represents the contribution of the *kth* HR dictionary atom to the final reconstruction. Since $\mathbf{A}^n$ is *jointly* sparse, it follows that the rows, $\boldsymbol{\alpha}_T^{n,k}$, contain either coefficients or zeros, within some predefined threshold.

Thus, *no reduction operator needs to be applied* to achieve a reduced system, as is suggested in [14][16].

b) The term , $\mathbf{E}_k^n$, is factorised according to a singular value decomposition:

$$\mathbf{E}_k^n = \mathbf{U}_k^n \mathbf{\Delta}_k^n \left(\mathbf{V}_k^n\right)^T \tag{7.4}$$

Note that the singular value, $\delta_{1,1}^{n,k}$, represents the most significant contributions to the decomposition.

c) The *most significant* sources of error in the term Equation (7.7) can be negated, and as such the term minimised, by setting the updated dictionary, $\hat{\mathbf{d}}_k^n$, as the first column of $\mathbf{U}_k^n$, denoted $\mathbf{u}_1^{n,k}$, and the sparse reconstruction coefficient, $\hat{\boldsymbol{\alpha}}_T^{n,k}$, as the first column of $\mathbf{V}_k^n$, denoted $\mathbf{v}_1^{n,k}$, multiplied by $\delta_{1,1}^{n,k}$:

$$\hat{\mathbf{d}}_k^n = \mathbf{u}_1^{n,k} \tag{7.8}$$

$$\hat{\boldsymbol{\alpha}}_T^k = \delta_{1,1}^{n,k} \mathbf{v}_1^{n,k} \tag{7.9}$$

This is applied through all N_{DP} dictionary elements, resulting in the updated HR dictionary, $\hat{\mathbf{D}}_h^n$. Note that to accelerate convergence, as proposed in [14], the most current $\hat{\mathbf{d}}_k^n$ and $\hat{\boldsymbol{\alpha}}_T^{n,k}$ values are updated online i.e the kth dictionary element and alpha coefficients are updated before computation of the $(k+1)th$ update.

d) The atoms of the updated HR dictionary, $\hat{\mathbf{D}}_h^n$, are downsampled to generate the updated LR dictionary, $\hat{\mathbf{D}}_l^n$, to be used in the next iteration J-SparseFI sparse coefficient update.

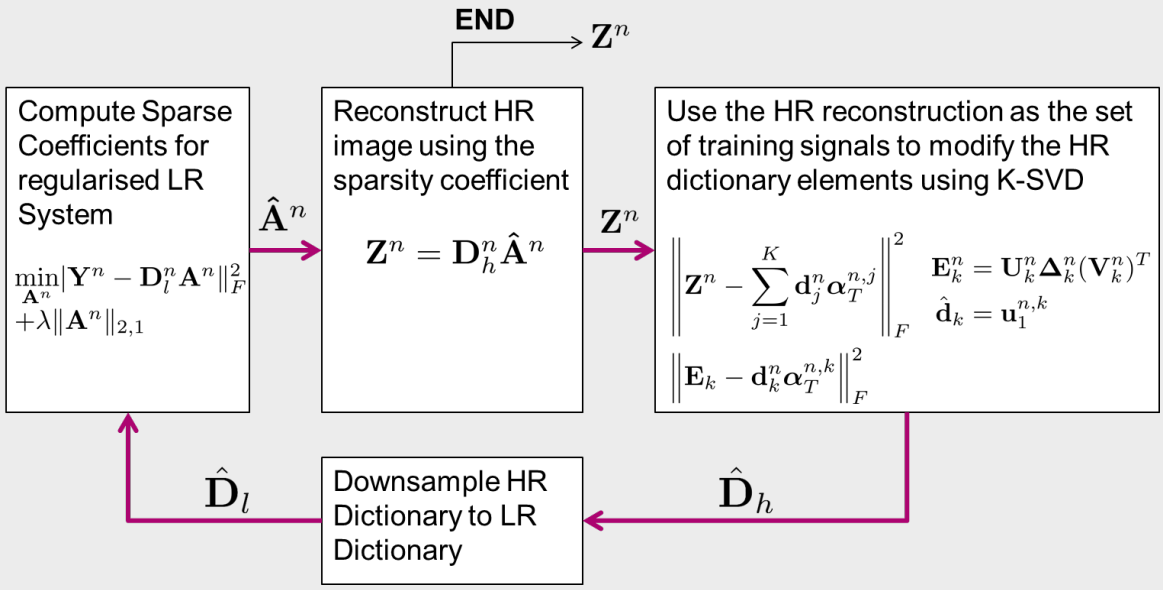
4. **Update Main Loop**

J = J + 1

Repeat until convergence within some predefined tolerance.

This K-SVD post training algorithmis illustrated in the flow diagram in Figure 7.3:

Figure 7.3: The K-SVD post training algorithm, detailed in Section 7.2, is depicted in flow diagram format.



This post training algorithm is tested for two different dictionary selection techniques in Chapter 8.

Chapter 8

K-SVD Dictionary Training Performance

In this chapter results for the K-SVD post training algorithm, detailed in Chapter 7, are presented for two different dictionaries. While on average K-SVD results in a small decrease in performance, in terms of image quality metrics, improvement is observed for the red edge WorldView-2 channel, channel 6. This improvement for channel six was consistent for two different dictionary selection methods.

8.1 Testing Overview

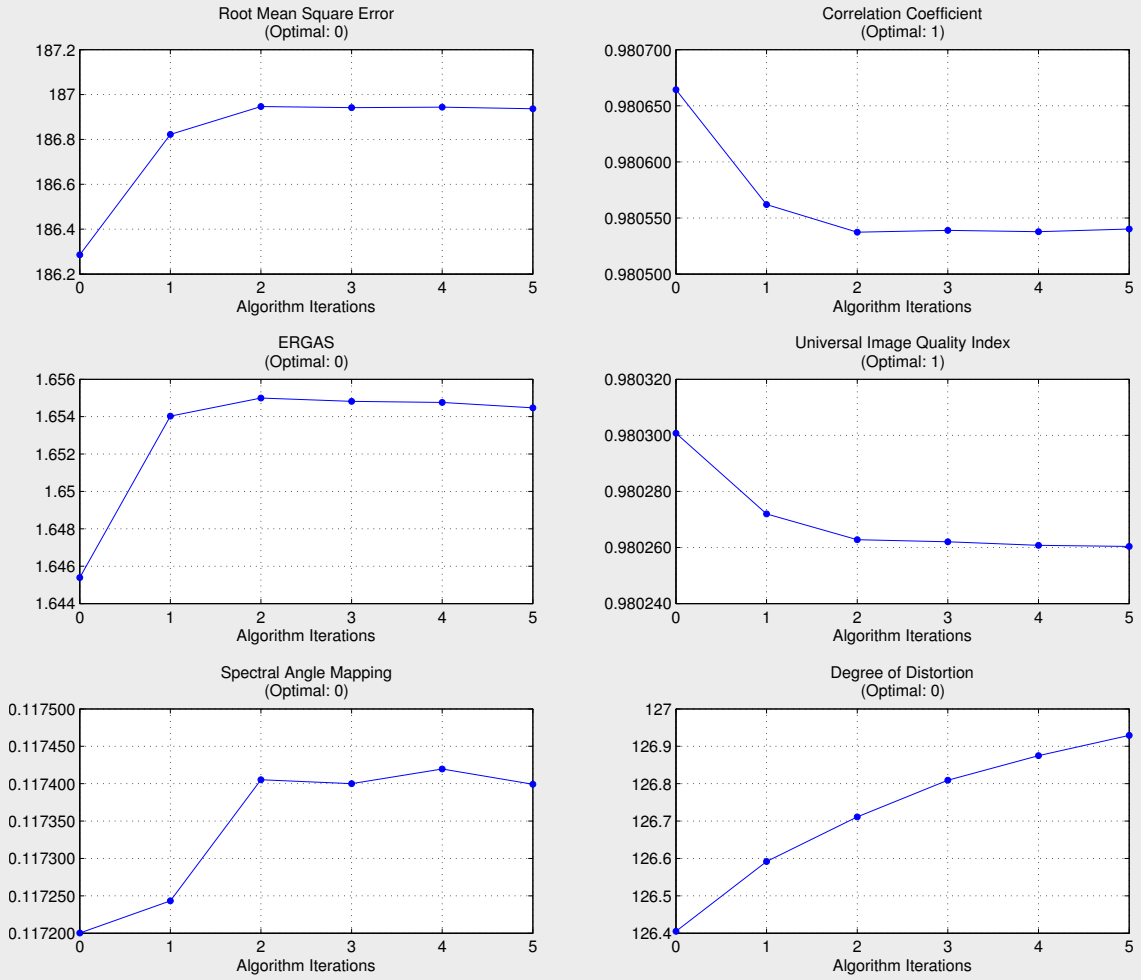
The synthetic test dataset was prepared as outlined in Chapter 4, using the same reduced reconstruction region to limit cpu usage, while still selecting dictionaries from a subset of the complete set of PanHR and PanLR image patches. Performance testing was conducted using a variable number of K-SVD iterations as implemented in the post training algorithm outlined in Chapter 7. For these tests, MS channels 2-6 were reconstructed, since benchtop testing revealed consistent improvement in channel 6.

K-SVD was applied to two separate dictionary selection methods, which both provide reasonable spatial performance, in order to assess the persistence of any underlying patterns across two completely different dictionaries. Both the benchmark NNP selection method and the anti-correlation based NL8 method were used to select dictionaries for K-SVD post training testing.

8.2 K-SVD Training for NNP Selected Dictionary

Up to 5 iterations of K-SVD was applied to an NNP selected dictionary of size $N_{DP} = 300$, a size that provides close to optimal performance for this particular test setup and dataset, according to the dictionary selection testing. The average image quality performance as a function of number of iterations of K-SVD is displayed in Figure 8.1.

Figure 8.1: K-SVD dictionary post training performance is evaluated for the standard image quality assessment metrics for various iterations, as detailed in Appendix A. Note that 0 iterations of K-SVD corresponds to the benchmark case where no K-SVD dictionary training is applied.

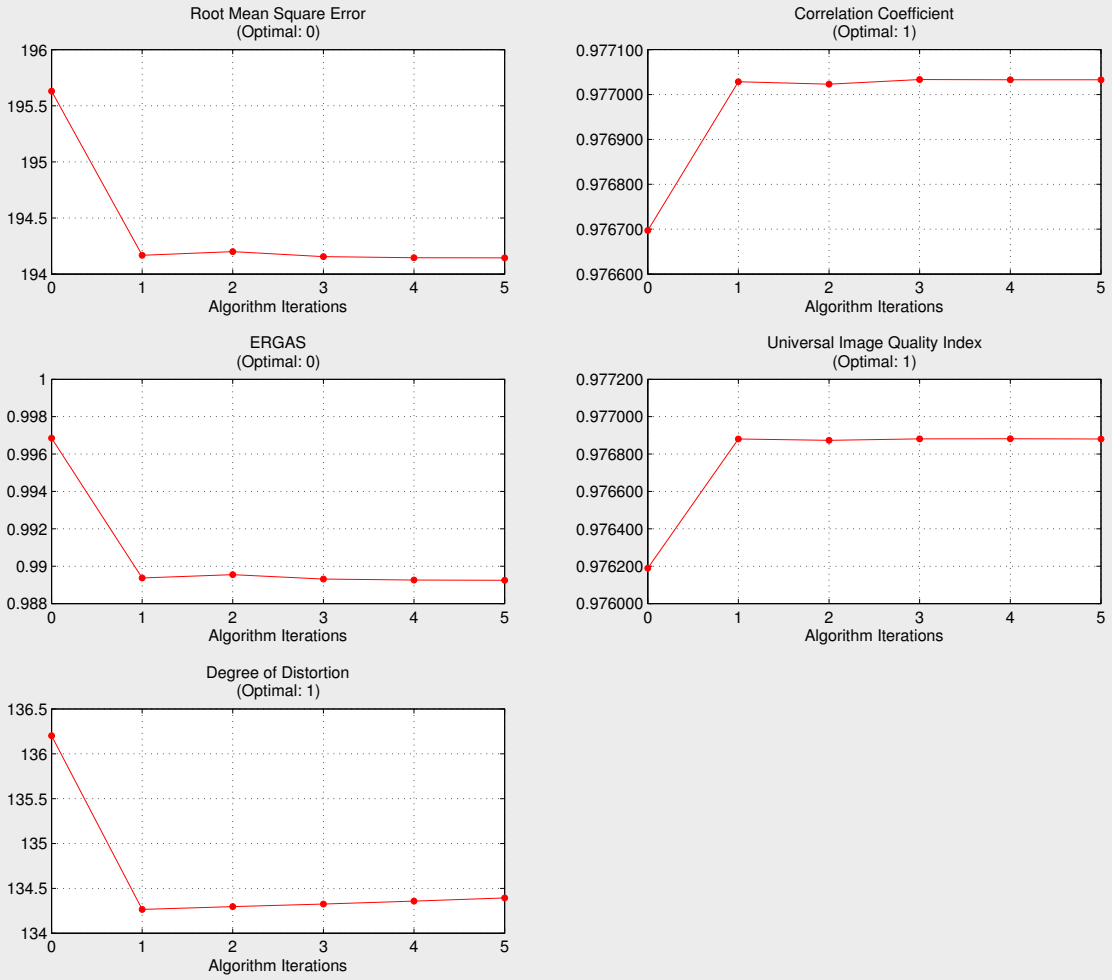


Note that average performance is observed to slightly decrease for all image quality metrics, after one iteration of K-SVD training. After this single iteration the image quality metrics, excluding degree of distortion, appear to converge on a relatively stable value. Note that the orders of magnitude of variation involved are much lower than was incurred by the dictionary

selection testing.

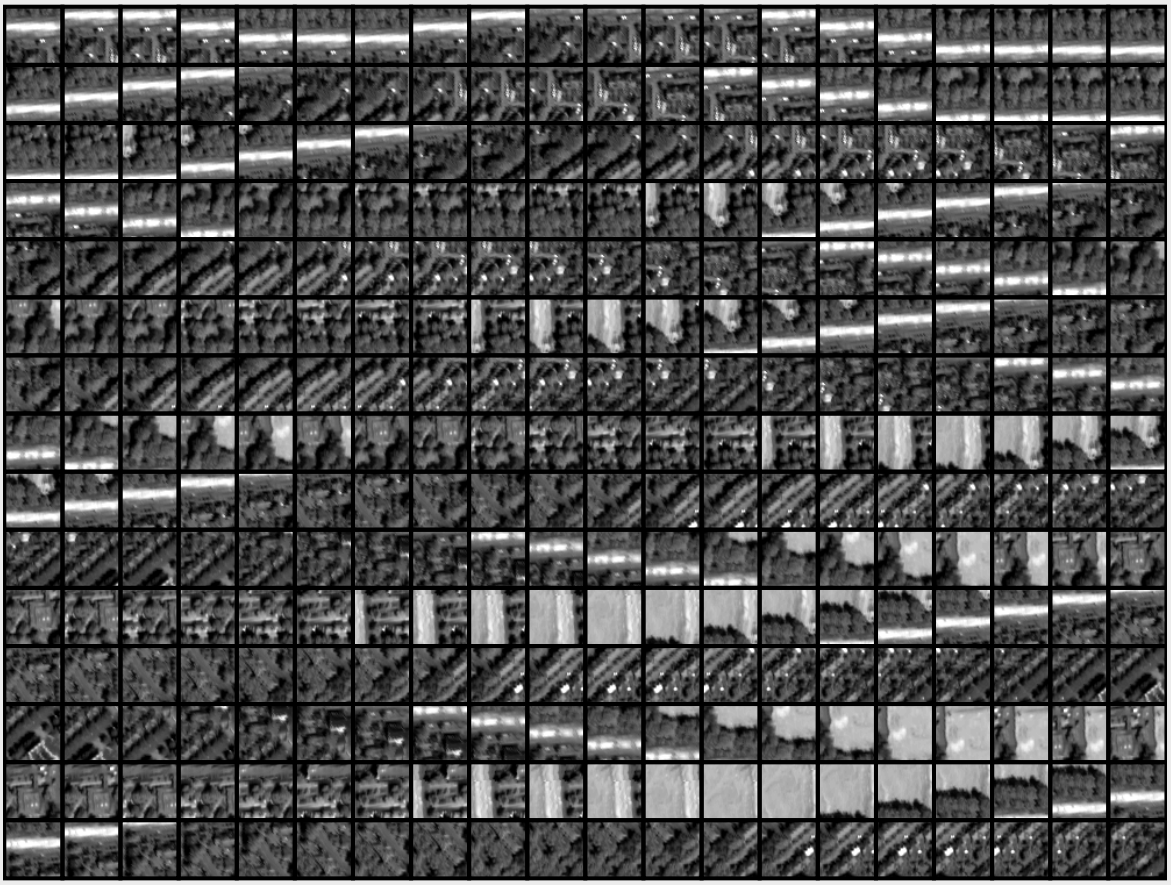
Another interesting trend identified in benchtop testing was the relative improvement of the image quality metrics observed for the red edge WorldView-2 channel 6, following one iteration of K-SVD dictionary training. This improvement is depicted below in Figure 8.2.

Figure 8.2: K-SVD dictionary post training performance of the red edge channel 6 is evaluated for the standard image quality assessment metrics for various iterations, as detailed in Appendix A. Note that 0 iterations of K-SVD corresponds to the benchmark case where no K-SVD dictionary training is applied.



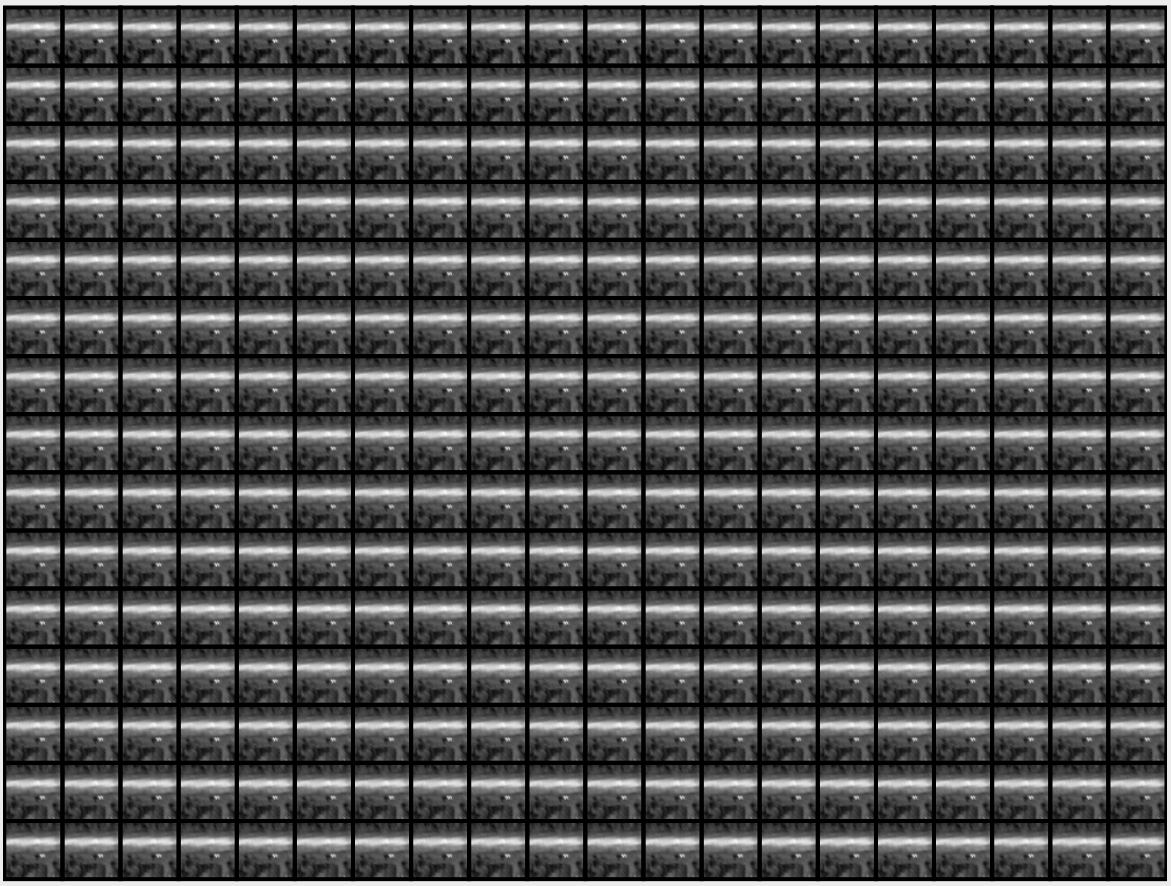
To assess the effect that K-SVD dictionary post training has on individual dictionary atoms, the complete HR dictionaries are graphed prior to K-SVD training and after one iteration of K-SVD training. The initial NNP selected dictionary of size $N_{DP} = 300$ is depicted in Figure 8.3.

Figure 8.3: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NNP selection method, prior to K-SVD dictionary training. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right.



The NNP selected dictionary after one iteration of K-SVD post training is depicted in Figure 8.4.

Figure 8.4: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NNP selection method, post one iteration of K-SVD dictionary training. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right.

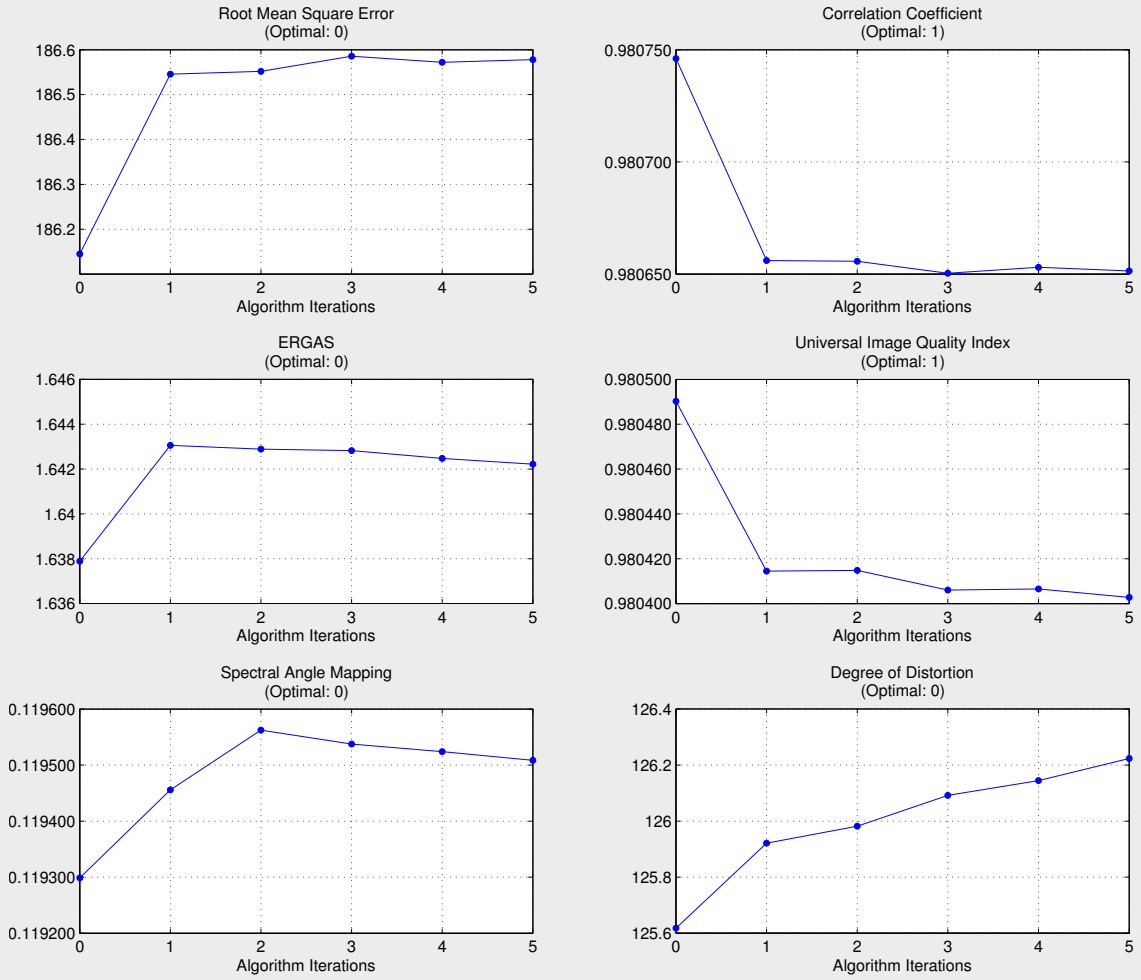


Note that K-SVD post training is observed to modify dictionary atoms to match the geometrical features prominent in the current patch under reconstruction. It is important to note, however, that the relative pixel intensities display variation from atom to atom i.e the dictionary is not composed of a single identical patch repeated for all atoms.

8.3 K-SVD Training for Anticorrelated (NL8) Selected Dictionary

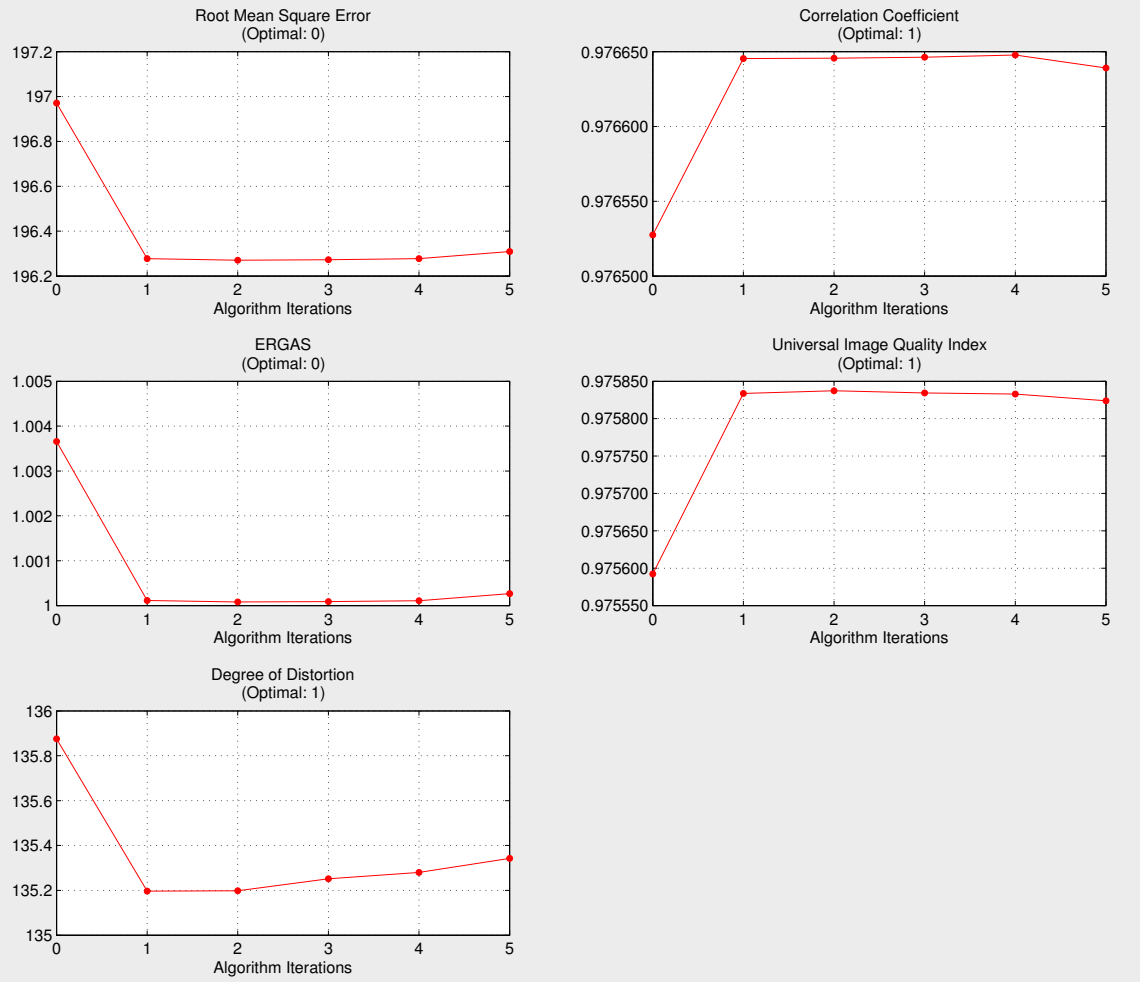
Up to 5 iterations of K-SVD was also applied to an NL8 anticorrelated dictionary of size $N_{DP} = 300$, a size that provides performance exceeding NNP selection for all assessed metrics, with the exception of SAM. The average image quality performance as a function of number of iterations of K-SVD is displayed in Figure 8.5.

Figure 8.5: K-SVD dictionary post training performance is evaluated for the standard image quality assessment metrics for various iterations, as detailed in Appendix A. Note that 0 iterations of K-SVD corresponds to the benchmark case where no K-SVD dictionary training is applied.



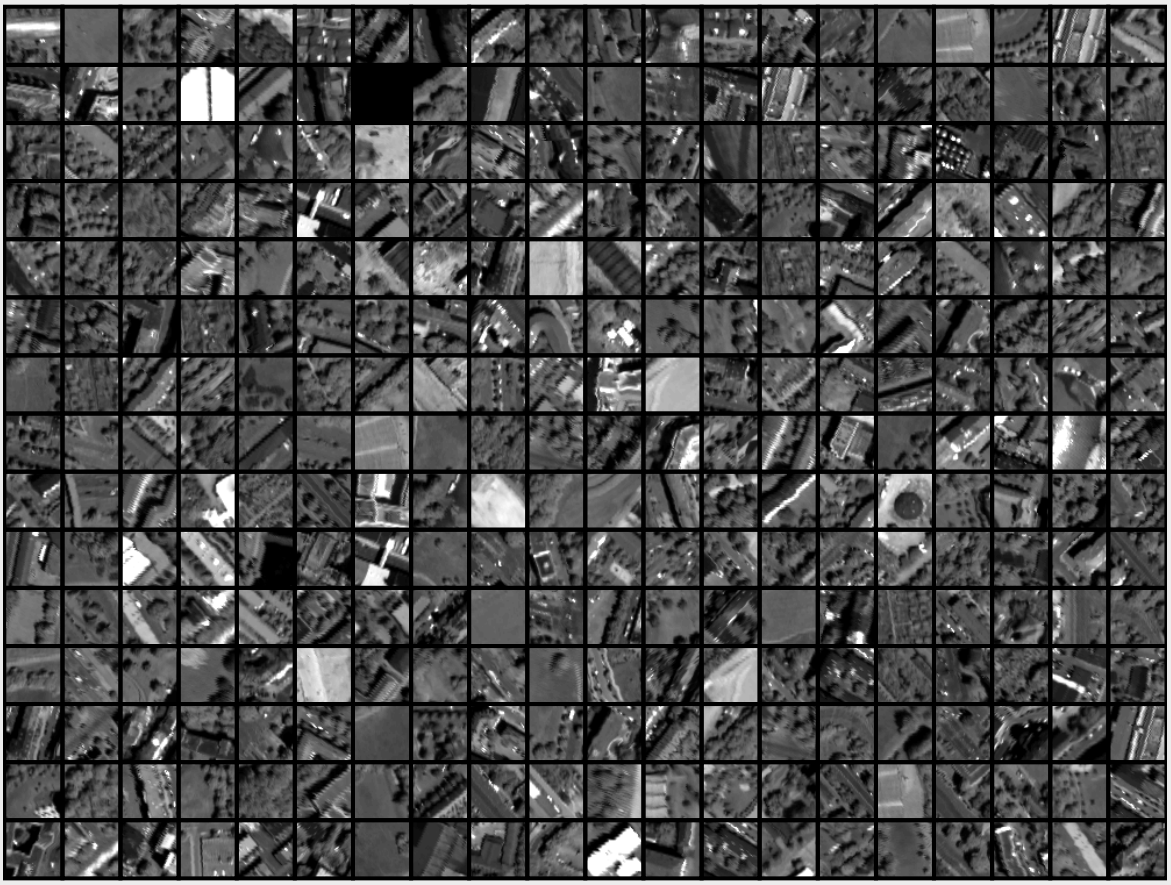
The improvement for the red edge channel 6 was also observed for NL8 dictionary selection, as is displayed below in Figure 8.6.

Figure 8.6: K-SVD dictionary post training performance of the red edge channel 6 is evaluated for the standard image quality assessment metrics for various iterations, as detailed in Appendix A. Note that 0 iterations of K-SVD corresponds to the benchmark case where no K-SVD dictionary training is applied.



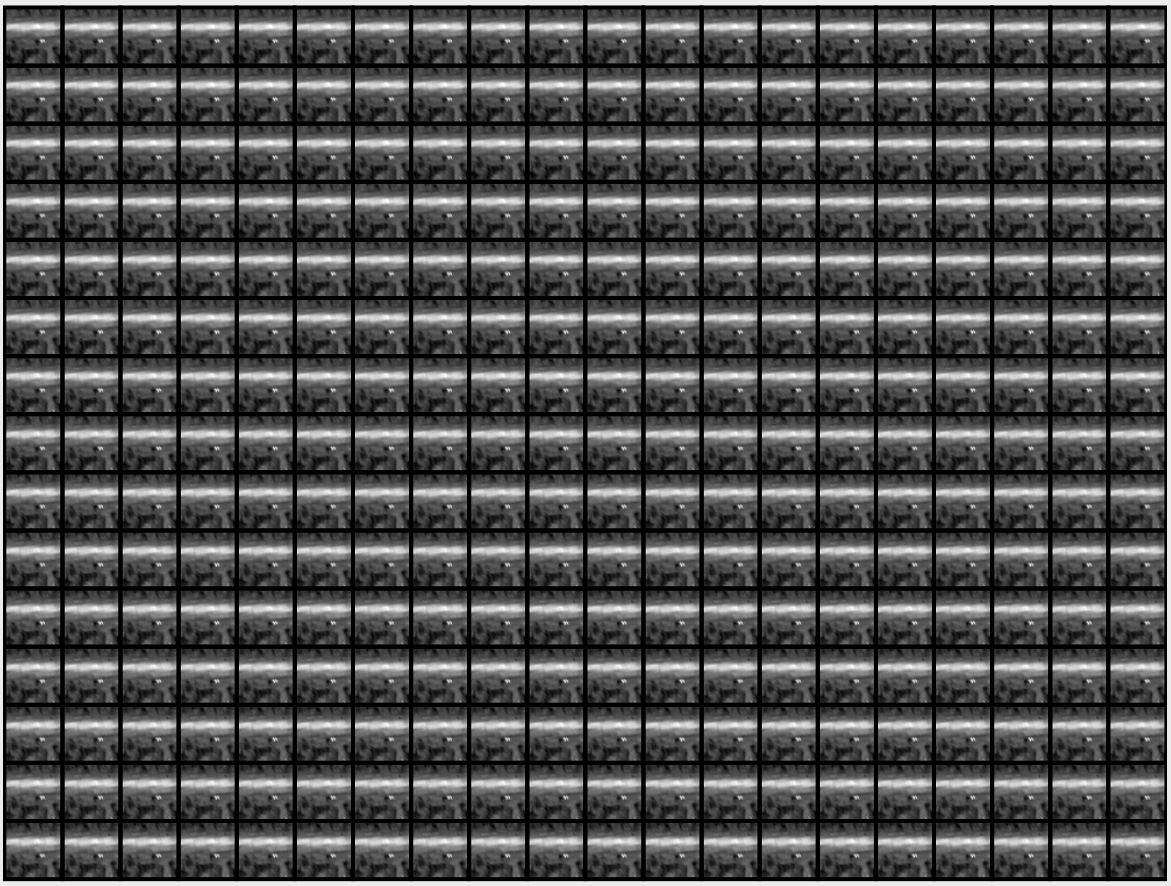
To assess the effect that K-SVD dictionary post training has on individual dictionary atoms, the complete HR dictionaries are graphed prior to K-SVD training and after one iteration of K-SVD training. The initial NL8 selected dictionary of size $N_{DP} = 300$ is depicted in Figure 8.7.

Figure 8.7: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NNP selection method, prior to K-SVD dictionary training. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right.



The NL8 selected dictionary after one iteration of K-SVD post training is depicted in Figure 8.8.

Figure 8.8: Example road patch HR dictionary, $\mathbf{D}_h^n$, for the NNP selection method, post one iteration of K-SVD dictionary training. The dictionary is displayed such that the first patch appears in the top left corner, with consecutive dictionary atoms read from left to right.

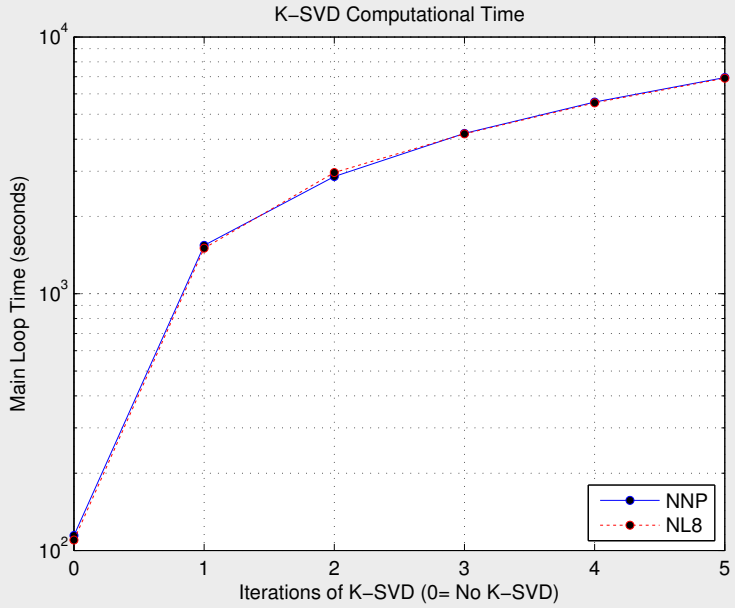


Note that as observed in NNP dictionary training, K-SVD post training also modified NL8 dictionary atoms to match the geometrical features prominent in the current patch under reconstruction, though as before, the relative pixel intensities display variation from atom to atom.

8.4 Computational Performance

It should be noted that although K-SVD post training had limited impact on the image reconstruction performance, the singular value decomposition of the 2500×5 $\mathbf{E}_k^n$ matrix was numerically expensive. This is illustrated in the computational time required for the main loop to complete, displayed in Figure 8.9.

Figure 8.9: The main loop time for 0-5 iterations of K-SVD post training on NNP and NL8 selected dictionaries, respectively.



The log y axis displays that the K-SVD post training represents and order of magnitude increase in computational time. Not apparent in Figure 8.9, due to the log y axis, is the linear increase in time with consecutive iterations of K-SVD.

Chapter 9

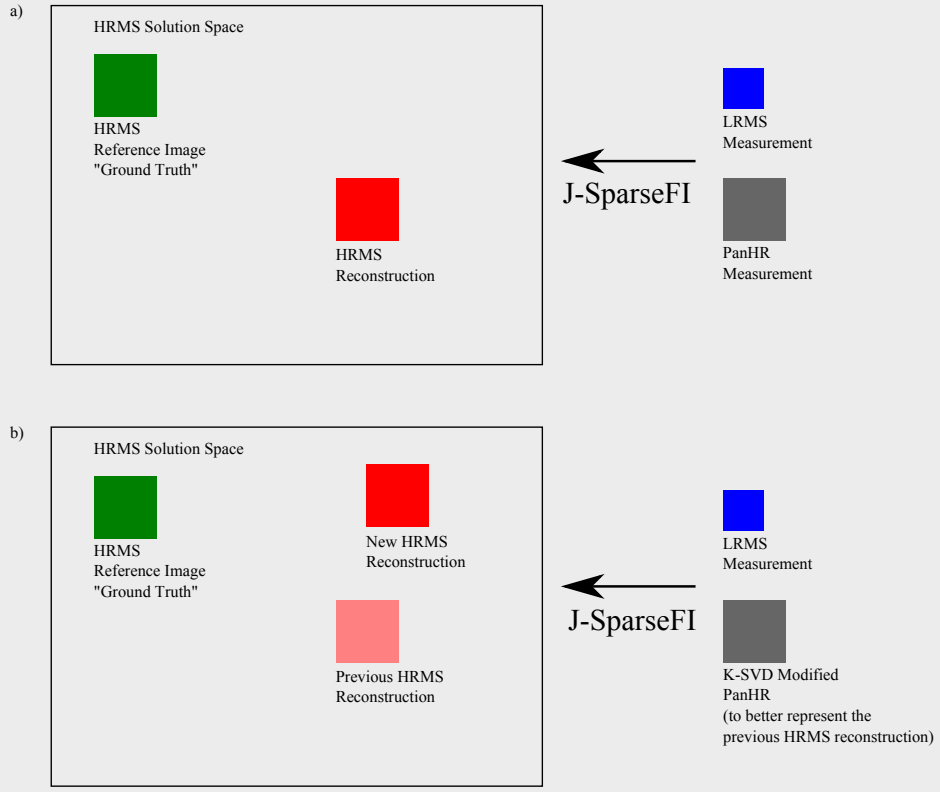
K-SVD Evaluation

Evaluation of the K-SVD results acquired in Chapter 8 is presented in this chapter. Improvements observed in the red edge WorldView-2 channel 6 are argued to be a result of this channels unique combination of poor LR correlation with the other sharpened channels, 2-5, and good LR correlation with the PanLR image. As such, K-SVD presents an opportunity for sharpening similar individual channels. A reposed form of the K-SVD algorithm, incorporating information from the LRMS measurement image is also incorporated and recommended for future investigation. Possible applications in Pan image denoising for J-SparseFI are also mentioned with regards to continued investigation.

9.1 Performance Evaluation

A central limitation of the K-SVD post training method implemented for J-SparseFI is the strong assumption that adapting the dictionary to better represent the *current iteration MSHR reconstruction*, $\mathbf{Z}^n$, will lead to an improvement in the next iteration reconstruction. Since the current reconstruction is likely, though not *necessarily*, representative of the HRMS ground truth, it is not immediately evident that adaptation of dictionary atoms will better represent the HRMS *ground truth*, which is the ultimate goal of any pan-sharpening method. This dilemma is illustrated in Figure 9.1.

Figure 9.1: a) Depicts the combination of the measurements using J-SparseFI to generate a HRMS reconstruction that lies somewhere in the HRMS solution space, relative to the HRMS reference image. In b) the PanHR measurement has been altered using K-SVD to better represent the previous HRMS reconstruction. However, this does not necessarily lead to a new HRMS reconstruction that is better representative of the HRMS reference image.



In fact this strong assumption is illustrated by considering that the training data is taken to be the current MSHR reconstruction, $\mathbf{Z}^n$. Thus initially:

$$\left\| \mathbf{Z}^n - \sum_{j=1}^{K} \mathbf{d}_j^n \boldsymbol{\alpha}_T^{n,j} \right\|_F^2 = \left\| \mathbf{E}_k^n - \mathbf{d}_k^n \boldsymbol{\alpha}_T^{n,k} \right\|_F^2 = 0 \tag{9.1}$$

Hence modification of the dictionary, such that the atoms represent the SVD rank one approximation of the error term, will lead to:

$$\left\| \mathbf{Z}^n - \sum_{j=1}^{K} \mathbf{d}_j^n \boldsymbol{\alpha}_T^{n,j} \right\|_F^2 = \left\| \mathbf{E}_k^n - \mathbf{d}_k^n \boldsymbol{\alpha}_T^{n,k} \right\|_F^2 \approx 0 \tag{9.2}$$

This is in fact what is observed in the K-SVD post training, with average image quality metrics slightly decreasing following the training iterations.

A suggestion for preserving a greater amount of spectral information implicit in the measured LR patch, $\mathbf{Y}^n$, involves the the minimization problem posed in Equation (7.1) being reposed as a system that combines the HR and LR dictionaries and reconstructions. The modified dictionary atoms used in the reposed problem are composed of a concatenation of the HR-Pan patches and the LR-Pan patches as follows:

$$\bar{\mathbf{D}}^n = \begin{bmatrix} \mathbf{D}_h^n \\ \beta \mathbf{D}_l^n \end{bmatrix} \tag{9.3}$$

Given that the HR dictionary elements contain considerably more entries than the LR dictionary elements, a weighting factor, β, is included to balance the representation of the LR measurement information in the dictionary training. The value of beta is set to approximately the square of the downsampling factor, as this represents the factor of difference between the number of entries in the LR and HR patches. Similarly, the modified reconstruction is formed from a concatenation of the HR and LR reconstructions:

$$\bar{\mathbf{Z}}^n = \begin{bmatrix} \mathbf{Z}^n \\ \beta \mathbf{Y}^n \end{bmatrix} \tag{9.4}$$

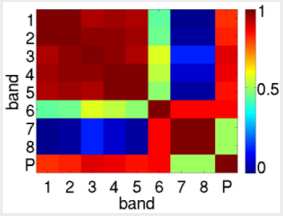
We thus pose the training problem, as in Equation (7.7), using the modified terms:

$$\left\| \bar{\mathbf{Z}}^n - \bar{\mathbf{D}}^n \mathbf{A}^n \right\|_F^2 = \left\| \bar{\mathbf{E}}_k^n - \bar{\mathbf{d}}_k \boldsymbol{\alpha}_T^{n,k} \right\|_F^2 \tag{9.5}$$

This reposed problem is solved in the manner afore described in the J-SparseFI with K-SVD Dictionary Training Algorithm. Due to the inclusion of the LR measurement information, it is postulated that this K-SVD implementation provides a more valid theoretical basis for dictionary post training. It is thus recommended for future investigation.

However it was also observed that K-SVD dictionary training provided modest improvement of the red edge channel 6. This improvement, while minor, was consistent for two differently selected coupled dictionaries. In assessing this result it is helpful to consult the WorldView-2 sensor correlation matrix, depicted in Figure 9.2, and spectral response function, as per Figure 4.2, for the individual MSLR channels.

Figure 9.2: The correlation matrix for the WorldView-2 MSLR images is shown below. High mutual correlation is graphically indicated by the red while low correlation approaches blue.



band	c	b	g	y	r	r-e	NIR-1	NIR-2	Pan
c	1	0.9948	0.95068	0.95388	0.94523	0.43563	-0.04678	-0.0453	0.8136
b	0.9948	1	0.97022	0.96797	0.95349	0.45932	-0.023	-0.021	0.83494
g	0.95068	0.97022	1	0.97208	0.94021	0.57295	0.11163	0.11368	0.8994
y	0.95388	0.96797	0.97208	1	0.99182	0.52701	0.026604	0.028747	0.87802
r	0.94523	0.95349	0.94021	0.99182	1	0.48402	-0.02784	-0.02552	0.84553
r-e	0.43563	0.45932	0.57295	0.52701	0.48402	1	0.85785	0.85656	0.86579
NIR-1	-0.04678	-0.023	0.11163	0.026604	-0.02784	0.85785	1	0.99889	0.49912
NIR-2	-0.0453	-0.021	0.11368	0.028747	-0.02552	0.85656	0.99889	1	0.49968
Pan	0.8136	0.83494	0.8994	0.87802	0.84553	0.86579	0.49912	0.49968	1

Note that channel 6 is the only channel to have low correlation to channels 2-5, but retain high correlation (> 0.8) with the PanLR channel. This correlation with the Panchromatic image is required since benchtop testing revealed that K-SVD did not improve channels 7 and 8, which maintain low correlation with channels 2-5 and the Pan image. Recall that the

$\mathbf{E}_k^n$ represents the reconstruction error resulting from the removal of the kth dictionary element. Thus the update of $\mathbf{d}_k^n$ and $\boldsymbol{\alpha}_T^{n,k}$ reduces the main source of error caused by removing $\mathbf{d}_k^n$. In other words the algorithm adapts $\mathbf{d}_k^n$ such that it better represents the reconstruction channel where it contributes the most.

The joint sparsity constraint likely favours dictionary atoms that jointly contribute to the reconstruction. Given the mutual correlation between channels 2-5, it is assumed that the majority of these atoms better represent these channels. It is also conceivable that certain dictionary atoms contribute strongly to the reconstruction of channel 6 and are simultaneously weighted less significantly in the reconstruction of channels 2-5. These atoms are, in a sense, more crucial to the reconstruction of channel 6. The likelihood of such circumstances increases when you consider the low mutual correlation between channel 6 and channels 2-5. Hence it is inferred that the modification of atoms that contribute strongly to channel 6 may be responsible for the modest performance increase observed in this channel, given their relative importance to the reconstruction of this decoupled channel.

Therefore by improving dictionary elements according to their greatest contribution, improvement of individual decoupled channels is conceivable. This would result in a weakening of the joint sparsity assumption, leading to a reduction in joint weightings, or alternatively, a reduction in the correlation between the magnitude of sparsity coefficients in different channels. In fact, this is observed for both the NNP and NL8 dictionaries, evident in their respective sparse reconstruction coefficients. Figure 9.3 displays the sparse reconstruction coefficients of the NL8 selected dictionary prior to K-SVD training.

Figure 9.3: The sparse reconstruction coefficients, α, are graphed for the reconstructed channels 2-6, for the NL8 selection method. Note that the magenta channel corresponds to the red edge channel 6.

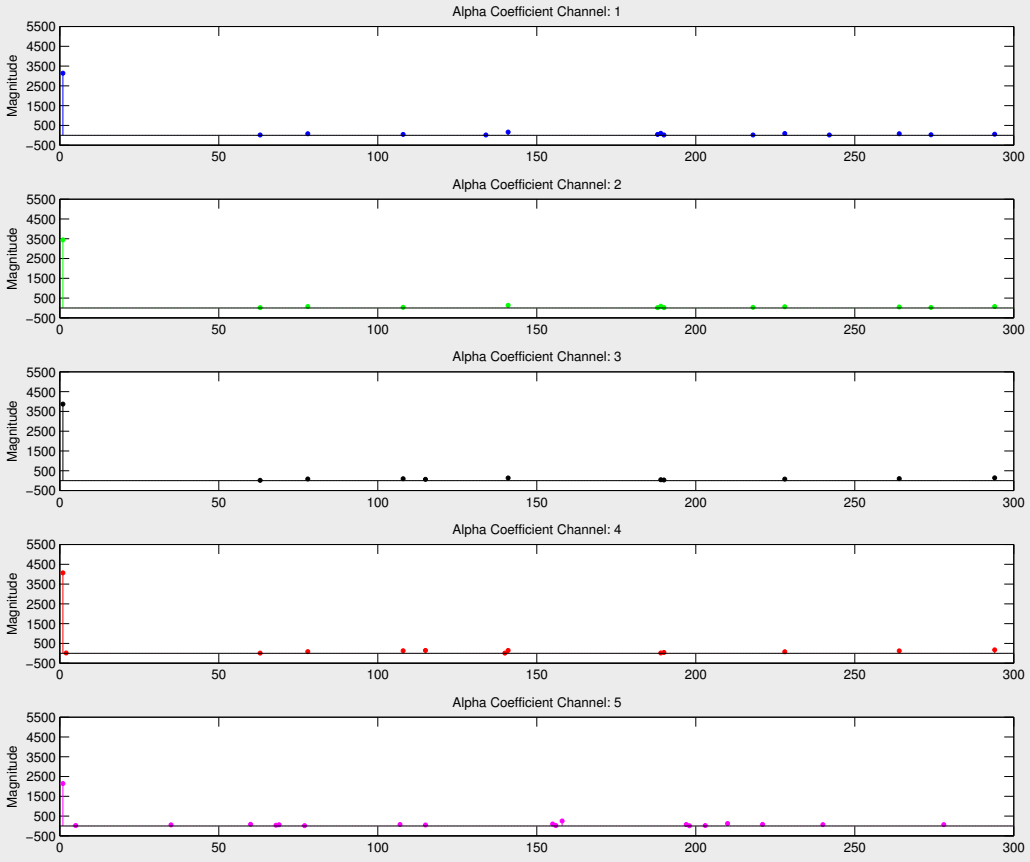
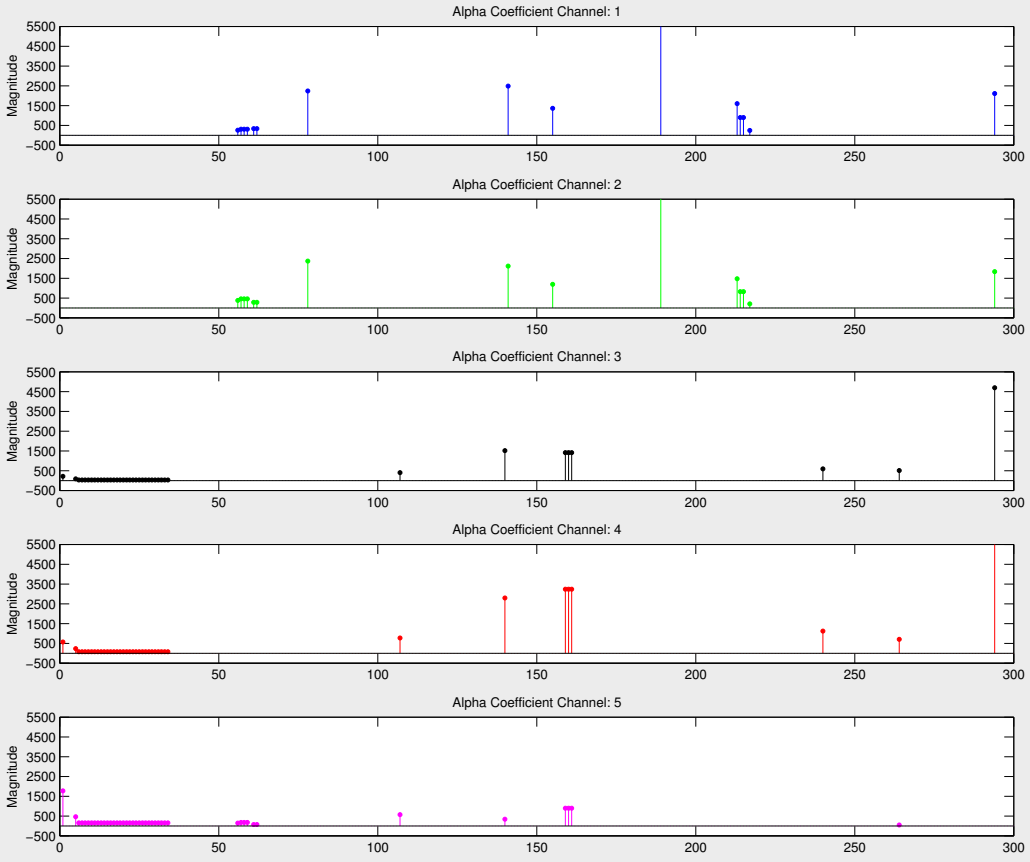


Figure 9.4 demonstrates the loss of joint sparsity in the sparse reconstruction coefficients, following one iteration of K-SVD.

Figure 9.4: The sparse reconstruction coefficients, α, are graphed for the reconstructed channels 2-6, for the NL8 selection method, after one iteration of K-SVD. Note that the magenta channel corresponds to the red edge channel 6.



Notable is the fact that the most heavily weighted atom is no longer the first atom in the dictionary, and furthermore, is not shared in all channels, demonstrating the effects of K-SVD training on the joint sparsity assumption. This trend is also evident in the NNP selected dictionaries, depicted in Appendix C. It should also be considered that distribution of alpha coefficients observed may also be the result of very similar LR dictionary atoms being passed to the solver. While this can account for the change in alpha coefficient distribution, as the solver experiences difficulty in selecting the most relevant patch, it does not necessarily ac-

count for the reduction in joint sparsity observed, particularly the reduction in joint sparsity for the most heavily weighted dictionary atom. It is argued that this is better explained by subtle differences in individual atoms that make them more suitable for respective channels in the reconstruction.

This lends support to the suggested explanation for the improvement of the decoupled red edge channel 6. However it is important to consider that this result was tested on only single data set with a single simulated sensor. In order to better verify both the characteristic sharpening of individual or decoupled channels, and the suggested mechanism of independent sharpening, additional testing on different simulated datasets, with synthetic sensors should be conducted. Specifically, a sensor should be constructed such that there exists a highly uncorrelated channel, that is well represented by the Pan channel, to test the assertion that K-SVD post training can improve channels with low correlation to the other MSLR channels, but high correlation to the Pan channel.

9.2 Conclusion and Recommendations

Given the strong assumption that the current reconstruction can be used as a training signal for the dictionary, present in the K-SVD post dictionary training algorithm investigated, it is evident that such approaches are limited in reconstruction improvement possibilities. Adapting the dictionary to better represent the *current iteration MSHR reconstruction*, $\mathbf{Z}^n$, assumes that the current reconstruction is likely, though not *necessarily* representative of the HRMS ground truth, and thus it is not necessarily valid that adaptation of dictionary atoms will better represent the HRMS *ground truth*, the ultimate goal of any pan-sharpening method.

However, a potential avenue of K-SVD improvement is identified in the sharpening of the red edge channel 6. While performance enhancement is modest, it provides evidence of plausible

improvement in channels which are well correlated to the Pan image but poorly correlated to the other channels being sharpened. An explanation for this relative performance increase is provided by the possibility of modifying dictionary atoms which are critical to the reconstruction of channel six, and mutually less important for channels 2-5. This effect is supported by the reduction in joint sparsity observed in the alpha coefficients post dictionary training.

In light of this analysis, the following recommendations are made:

- Additional investigation into the sharpening of individual decoupled channels should be conducted, to substantiate the results observed in this investigation. Ideally this would consist of replicated trials on an alternate data set, with a different simulated multispectral sensor.

- K-SVD trials should be conducted using noisy HR and LR Pan images. Given the observed alteration of dictionary elements to match the geometry of the current patch under reconstruction, it is postulated that K-SVD may be able to extract underlying features from noisy data. In this capacity, K-SVD post training operations may be used to increase the robustness of J-SparseFI.

- The dictionary selection has approximately an order of magnitude greater effect on the performance of J-SparseFI than the K-SVD dictionary training. Furthermore, K-SVD post training was observed to increase computational times by an order of magnitude. As such, further research efforts in performance enhancement should focus on dictionary selection rather than the proposed K-SVD algorithm.

Conclusion and Outlook

A summary of the most significant findings of this investigation is presented. Recommendations regarding both dictionary selection and dictionary training for J-SparseFI are outlined. An outlook toward future investigations in both dictionary selection and training is presented. The significance of this work is placed in the broader context of J-SparseFI hyperspectral and multispectral image fusion, an extension of the pan-sharpening problem.

Firstly, the goals and deliverables of this work, as outlined in Section 2.2, are itemised as follows:

- **Software Deliverables:** 10 coupled dictionary selection methods were implemented in the high performance J-SparseFI c++ application. A K-SVD post training algorithm was proposed and implemented. All software was run on the Leibniz-Rechenzentrum (LRZ) superMUC supercomputing facilities, Garching, Munich.

- **Testing:** The 10 dictionary selection methods were tested for dictionaries of sizes $50 \leq N_{DP} \leq 5000$ on synthetic WorldView-2 MSLR and PanHR data, synthesised from airborne HySpex datacube acquired over Munich, Germany. Testing was also performed for the proposed K-SVD post training algorithm, for training iterations up to 5. All testing was conducted using the SuperMUC supercomputing facilities.

- **Recommendations:** An in depth analysis of the proposed dictionary selection methods and their respective performance was undertaken. Three dictionary selection methods have been recommended on the basis of reconstruction spatial accuracy, robustness spectral fidelity. Future work regarding coupled dictionary selection and similar studies for J-SparseFI HSMS image fusion have been proposed. Analysis and critical

evaluation of the K-SVD post training algorithm was also conducted. Possible future investigations include a reposed K-SVD problem, incorporating information from the LR measurement image and possible denoising applications.

The main outcome of this work is the characterisation of the effects of different dictionary selection methods on the performance of the J-SparseFI algorithm for Pan-sharpening problems. A complex tradeoff between spectral fidelity and spatial accuracy was uncovered when considering dictionary selection methods for J-SparseFI. Central to the outcome of this work is the identification of similarity based local coupled dictionary selection performing better in spectral fidelity metrics and probabilistically dissimilar dictionary selection performing better in spatially weighted metrics. It should be noted that in all cases, inclusion of the HR and LR Pan patches corresponding to the patch undergoing reconstruction was *mandatory* for achieving competitive reconstruction performance.

These performance differences were explained in terms of the relative weightings assigned to dictionary atoms in the LR system by the sparse reconstruction coefficients, corresponding to performance variation in the HR system. In the case of similarity base dictionary selection a more even distribution of sparse reconstruction coefficient magnitude is observed, allowing for more pixel based refinement, improving spectral fidelity but increasing the magnitude of acumulated errors. In the case of probabilistically dissimilar selection methods, the prominent weighting of the first dictionary atom, corresponding to the patch under reconstruction, results in enhanced preservation of scene geometrical information but conversely permits less pixel based refinement, reducing spectral fidelity.

Specifically, three recommendations for dictionary selection are made, base on the criteria of spatial performance, robustness and spectral fidelity:

- **Spatial Performance:** The random dictionary selection method (NL9) yields the highest performance in all metrics with the exception of SAM. Random dictionary

selection also provides an improvement in the SAM with respect to the current NNP selection benchmark.

- **Robustness:** The anti-correlated dictionary selection method (NL8) demonstrates consistently decent image quality metrics, with the exception of SAM, for a wide range of dictionary sizes, N_{DP}. Improvement over the benchmark NNP selection is observed for all metrics excluding SAM.
- **Spectral Performance:** The patch angle mapping dictionary selection method (NL7) demonstrates outstanding spectral fidelity, with the lowest average SAM values. Furthermore, impacts to spatially weighted metrics are a minimised, given the HR similarity comparison based selection.

An avenue for continued investigation is the possibility of combined coupled dictionary selection, with selection based on two or more selection methods. In this capacity it is plausible to imagine a properly weighted combination promoting performance in both spatial and spectral metrics. As such, coupled dictionary selection based on a combination of metrics is recommended for future investigation.

The pan-sharpening performance results of the coupled dictionary selection investigation are of broader significance to the problem of hyperspectral-multispectral image fusion using the J-SparseFI algorithm. Although performance is reported to be excellent in spatially weighted image quality metrics, J-SparseFI does not excel at spectral fidelity for hyperspectral-multispectral image fusion [17]. Given that the J-SparseFI approach to hyperspectral-multispectral image fusion is, in simplified terms, to consider bunches of hyperspectral channels sharpened by their respective multispectral channels as a series of pan-sharpening problems, the conclusions drawn regarding dictionary selection for pan-sharpening likely remain valid for hyperspectral-multispectral image fusion. Thus, there exists the possibility of improving the spectral performance of J-SparseFI hyperspectral-multispectral image fusion using similarity based coupleddictionary selection. Hence, an in-depth investigation into

the effects of dictionary selection for J-SparseFI hyperspectral-multispectral image fusion is *highly recommended*.

Given the strong assumption that the current reconstruction can be used as a training signal for the dictionary, present in the K-SVD post dictionary training algorithm investigated, it is evident that such approaches are limited in reconstruction improvement possibilities. Adapting the dictionary to better represent the *current iteration MSHR reconstruction*, $\mathbf{Z}^n$, assumes that the current reconstruction is likely, though not *necessarily* representative of the HRMS ground truth, and thus it is not necessarily valid that adaptation of dictionary atoms will better represent the HRMS *ground truth*, the ultimate goal of any pan-sharpening method.

However, a potential avenue of K-SVD improvement is identified in the sharpening of the red edge channel 6. While performance enhancement is modest, it provides evidence of plausible improvement in channels which are well correlated to the Pan image but poorly correlated to the other channels being sharpened. An explanation for this relative performance increase is provided by the possibility of modifying dictionary atoms which are critical to the reconstruction of channel six, and mutually less important for channels 2-5. This effect is supported by the reduction in joint sparsity observed in the alpha coefficients post dictionary training.

Additional trials are recommended to further validate the mechanism of operation and implementation described. This initial K-SVD post training approach, may be improved by reposing the problem and including additional measurement information, as described in Chapter 9. However given the modest performance difference observed for K-SVD post training in this investigation, such efforts are likely better spent investigating dictionary selection, given the order of magnitude greater effect on overall performance.

A further possibility for the application of K-SVD dictionary training include the possibility for robust reconstruction given noisy Pan data. Given the adaptation of dictionary atoms to cancel the greatest source of error they contribute to the current reconstruction, the possibility for dictionary noise reduction and subsequent reconstruction robustness exists. However, this speculation requires considerable testing, in order to conclude whether this application is valid and suitable for the current implementation. Thus dictionary denoising should be investigated with regards to the K-SVD post training operation.

The dictionary selection and training assessments made in this investigation yield interesting and subtle insight into the J-SparseFI Pan-sharpening algorithm being developed at the DLR, and are of broad relevance to the topic of sparse reconstruction and compressive sensing. Performance insight and possible improvements to J-SparseFI is significant, given that this algorithm is currently a state of the art method. It is hoped that continued investigation into these areas, particularly dictionary selection, yields further performance and insight for J-SparseFI pan-shaprening, hyperspectral-multispectral image fusion and the field of compressive sensing.

Appendix A

Analysis of Experimental Results

A.1 Image Quality Metrics

The performance of the dictionary learning methods implemented is quantitatively evaluated in accordance with the following frequently used metrics.

A.1.1 Root Mean Square Error (RMSE)

The root mean square error (RMSE) provides a comparison between the original image and the pan-sharpened image by directly calculating the vectorlength change in pixel values. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X_{i,j} - \hat{X}_{i,j}\right)^2} \tag{A.1}$$

Where $X_{i,j}$ is the pixel value of the original image $\mathbf{X}$ and $\hat{X}_{i,j}$ is the pixel value of the pan sharpened image $\hat{\mathbf{X}}$. The relative error between the images is minimized as the RMSE becomes smaller.

A.1.2 Correlation Coefficient (ρ)

The Pearson correlation coefficient measures the linear correlation between sets of variables. In the context of image processing, this provides a measure of the similarity of spectral fea-

tures within corresponding images. It is defined as the covariance between the two variables divided by the product of their standard deviations:

$$\rho = \frac{\sum_{i,j}\left[(X_{i,j} - \bar{x}) \cdot \left(\hat{X}_{i,j} - \bar{\hat{x}}\right)\right]}{\sqrt{\left[\sum_{i,j}(X_{i,j} - \bar{x})^2\right] \cdot \left[\sum_{i,j}\left(\hat{X}_{i,j} - \bar{\hat{x}}\right)^2\right]}} \tag{A.2}$$

Where $\bar{x}$ and $\bar{\hat{x}}$ are mean values of the original image $\mathbf{X}$ and the pan-sharpened image $\hat{\mathbf{X}}$, respectively. High positive correlation is indicated by a ρ value approaching $+1$, while a value approaching 0 indicates no correlation between the image sets.

A.1.3 Degree of Distortion (D)

The degree of distortion reflects the average magnitude of the error between two images. It is defined as:

$$\mathrm{D} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|X_{i,j} - \hat{X}_{i,j}| \tag{A.3}$$

Optimally the distortion in the pan-sharpened image is small while the value of D is small.

A.1.4 The Universal Image Quality Index (Q-Average)

The Universal Image Quality Index (UIQI) is a frequently used metric for assessment of the quality of image sharpening. It combines three factors: Loss of correlation, luminance distortion and contrast distortion. It is defined as follows:

$$Q_0 = \frac{\sigma_{x\bar{\hat{x}}}}{\sigma_x \sigma_{\bar{\hat{x}}}} \times \frac{2\bar{x}\bar{\hat{x}}}{\left(\bar{x}^2 + \bar{\hat{x}}^2\right)} \times \frac{2\sigma_x \sigma_{\bar{\hat{x}}}}{(\sigma_x^2 + \sigma_{\bar{\hat{x}}^2})} \tag{A.4}$$

The optimal value of Q-average is 1.

A.1.5 Error Relative Dimensionless Global Error in Synthesis (ERGAS)

The Error Relative Dimensionless Global Error in Synthesis (ERGAS) describes the overall quality of the reconstructed image. It is defined as.

$$\text{ERGAS} = 100\frac{h}{l}\sqrt{\frac{1}{N_Y}\sum_{k=1}^{N_Y}\left[\frac{\text{RMSE}(I_Z(k))}{mean(I_Z(k))}(k)\right]^2} \tag{A.5}$$

Where h/l is the ratio between the pixel sizes of the panchromatic and original multispectral images, $\text{RMSE}(I_Z(k))$ and $mean(I_Z(k))$ are the root mean square and mean values of the kth band, respectively. A small ERGAS value is desirable as it indicates small spectral distortion, indicating good preservation of spectral information.

A.1.6 Spectral Angle Mapper (SAM)

The Spectral Angle Mapper (SAM) provides a comparison between image spectra. In computing this method, respective vectors are constructed in a space with the dimensionality of the number of spectral bands. The formula determines the spectral similarity between two images by calculating the *spectral* angle between them. Since only the *direction* of the spectra is used, the method is insensitive to unknown gain, and all illuminations are treated equally. Formally, SAM is invariant under scalar multiplication.

Given a reference vector $\mathbf{r}$ and an unknown spectrum $\mathbf{x}$ in N dimensional space, equal to the number of spectral bands, SAM is computed as:

$$\text{SAM} = \arccos\left(\frac{\mathbf{x}^T\mathbf{r}}{\|\mathbf{x}\|\|\mathbf{r}\|}\right) \tag{A.6}$$

The spectral difference between two images is minimised when the spectral angle between the two images is equal to zero.

Appendix B

Supplementary Results: Dictionary Selection

B.1 Reconstructions

Figure B.1: NNP Reconstruction

Figure B.2: NL1 Reconstruction

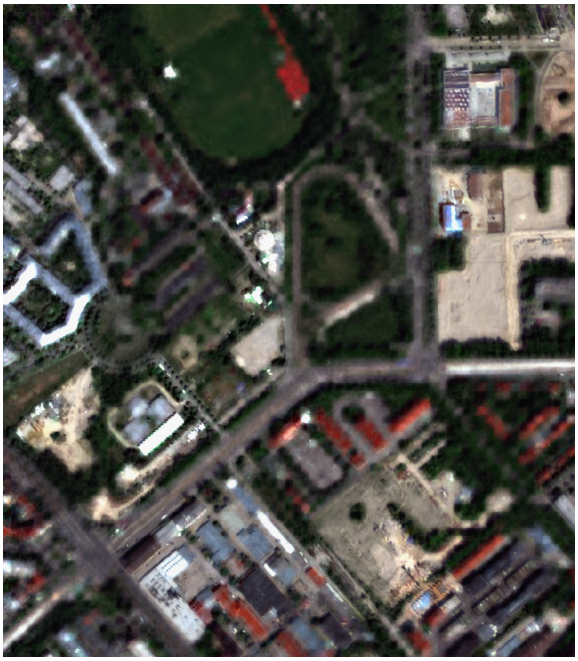


Figure B.3: NL2 Reconstruction

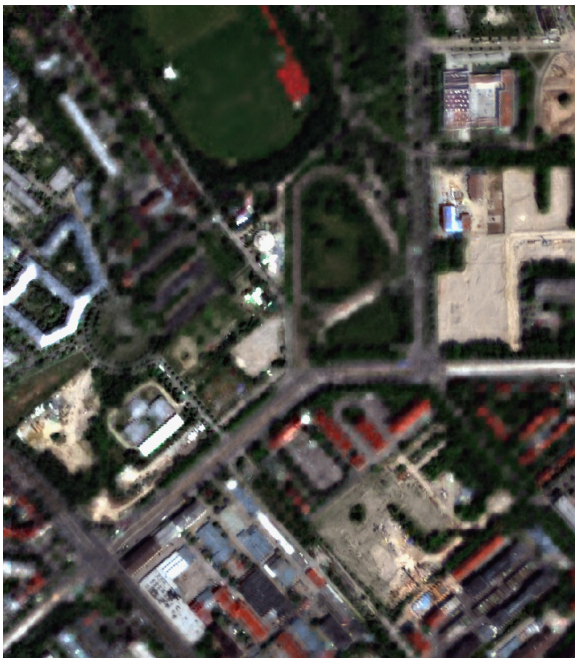


Figure B.4: NL3 Reconstruction

Figure B.5: NL4 Reconstruction

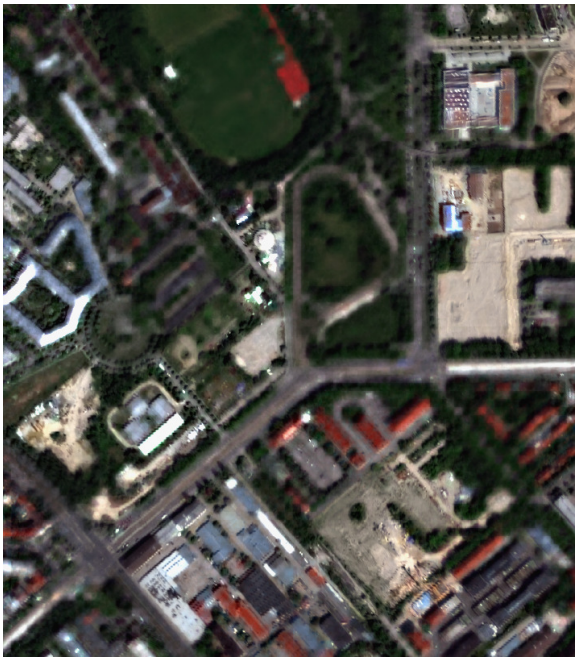


Figure B.6: NL5 Reconstruction

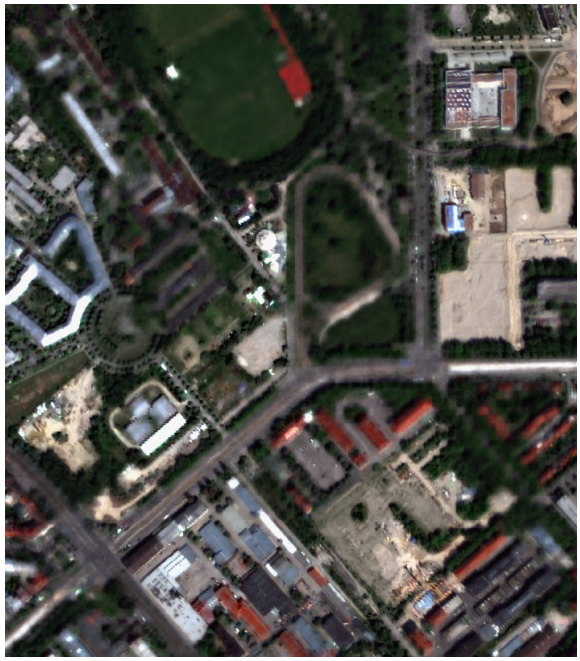


Figure B.7: NL6 Reconstruction

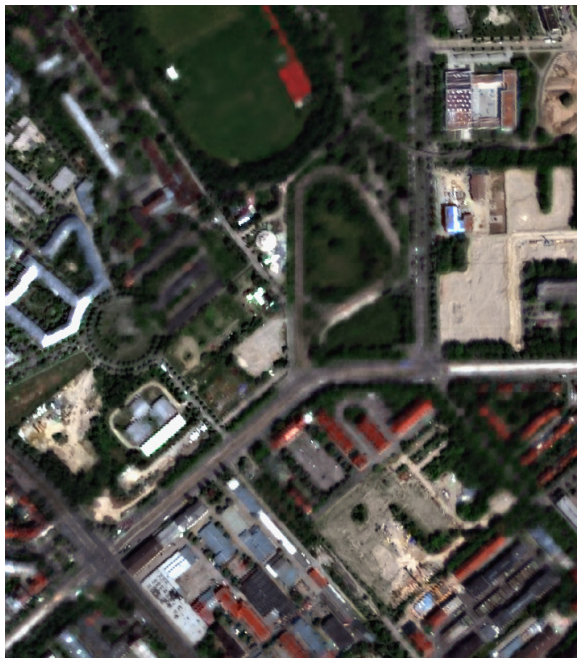


Figure B.8: NL7 Reconstruction

Figure B.9: NL8 Reconstruction

Figure B.10: NL9 Reconstruction

Figure B.11: NL10 Reconstruction

B.2 Alpha Coefficients

This section contains the alpha coefficients for the example road patch assessed in the thesis:

B.2.1 NNP Alpha Coefficients

Figure B.12: NNP Alphas

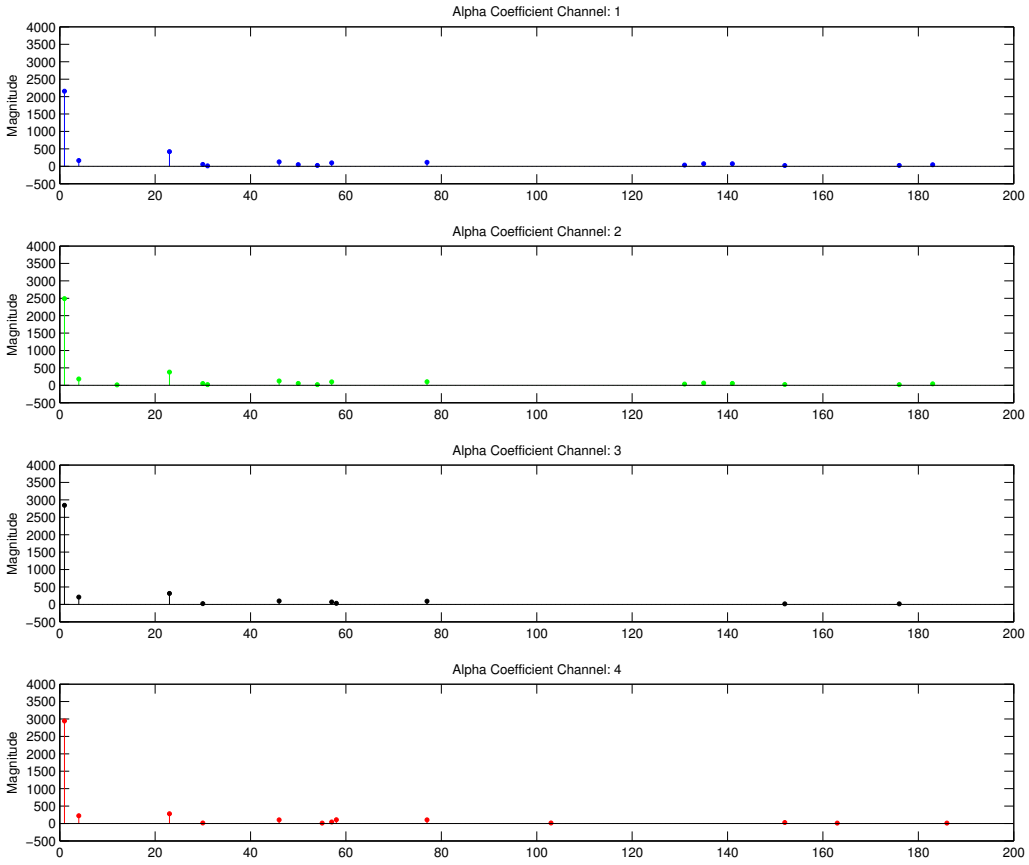


Figure B.13: NL3 Alphas

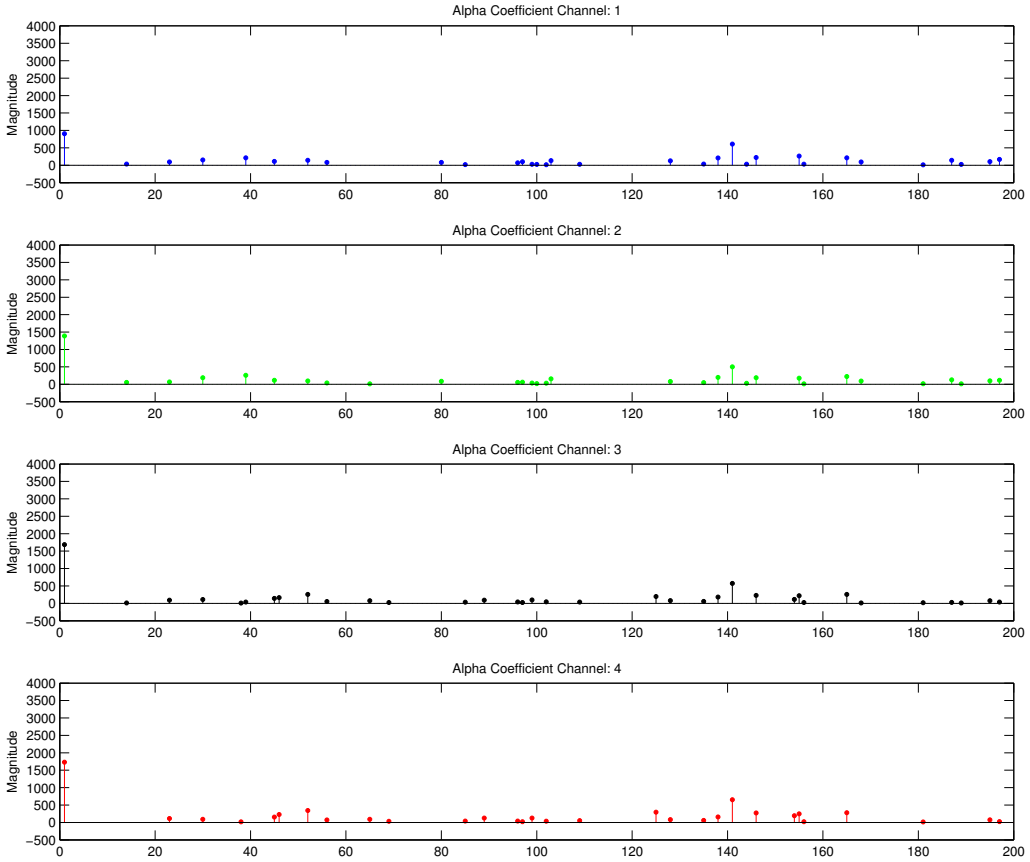


Figure B.14: NL4 Alphas

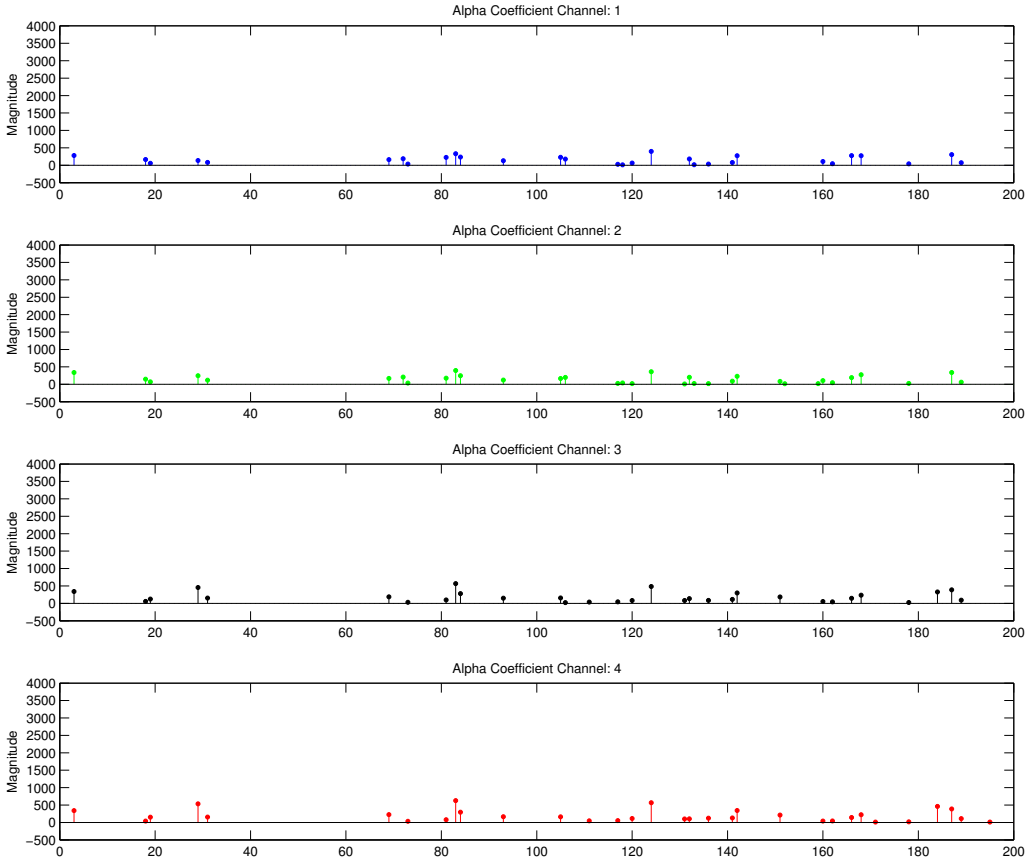


Figure B.15: NL5 Alphas

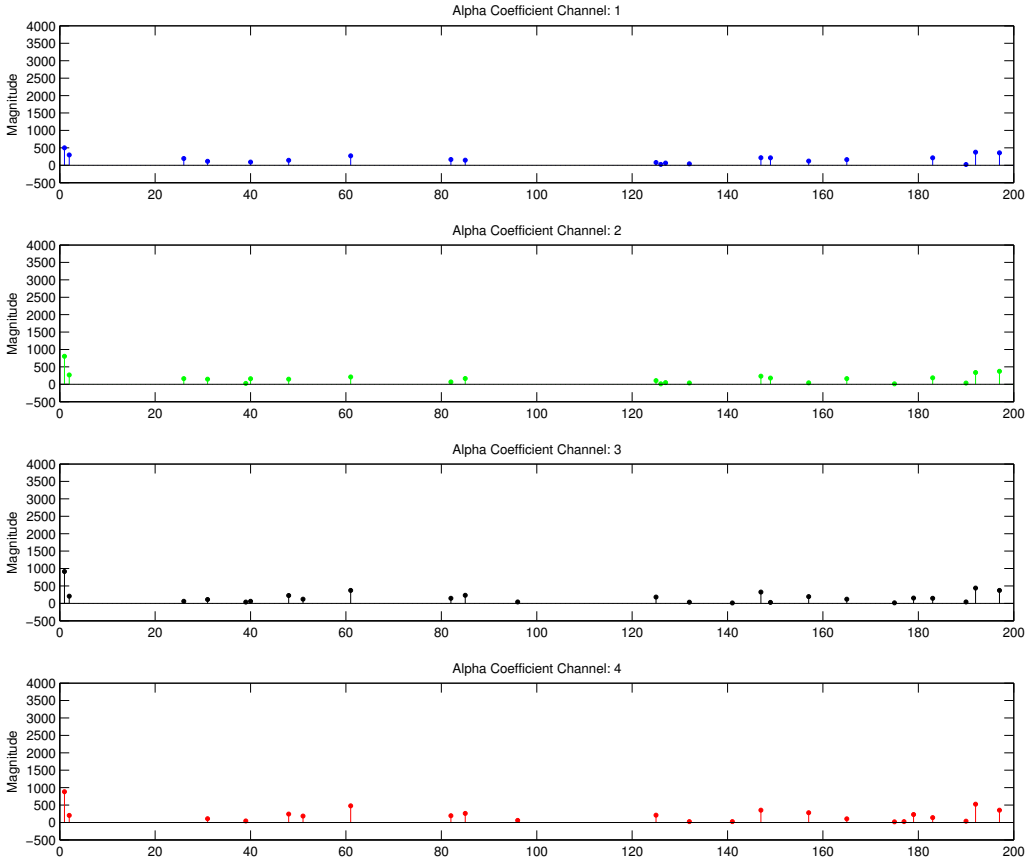


Figure B.16: NL6 Alphas

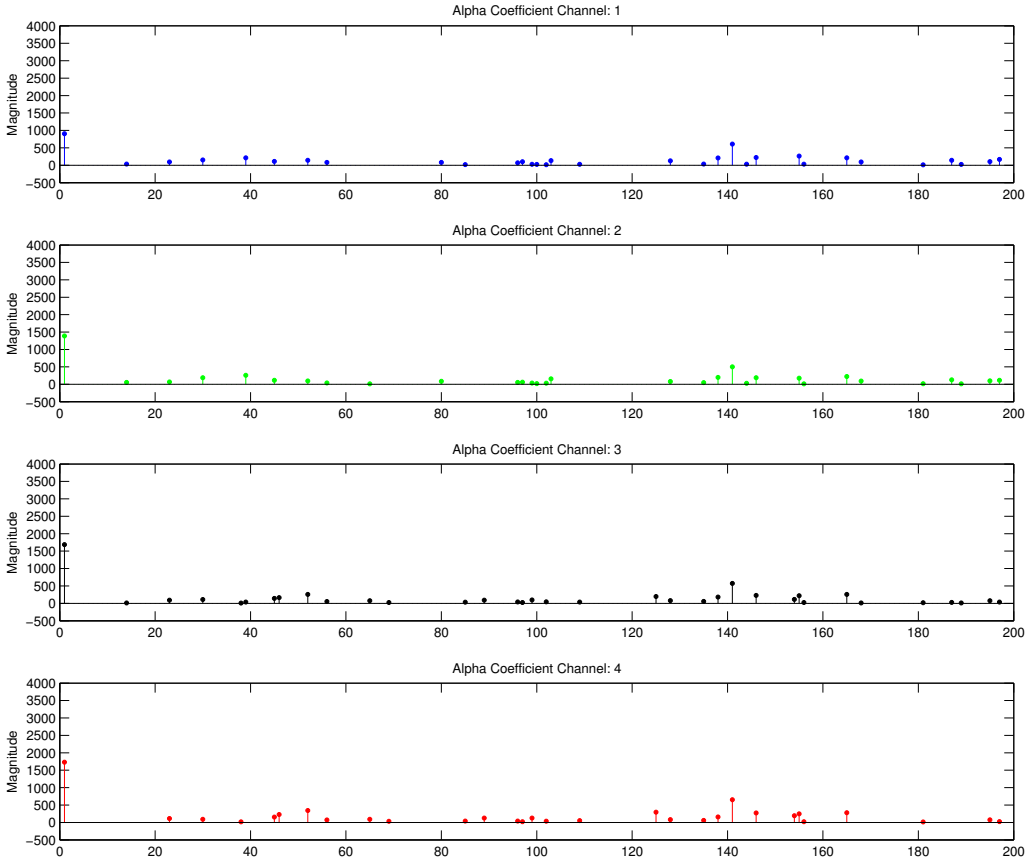


Figure B.17: NL7 Alphas

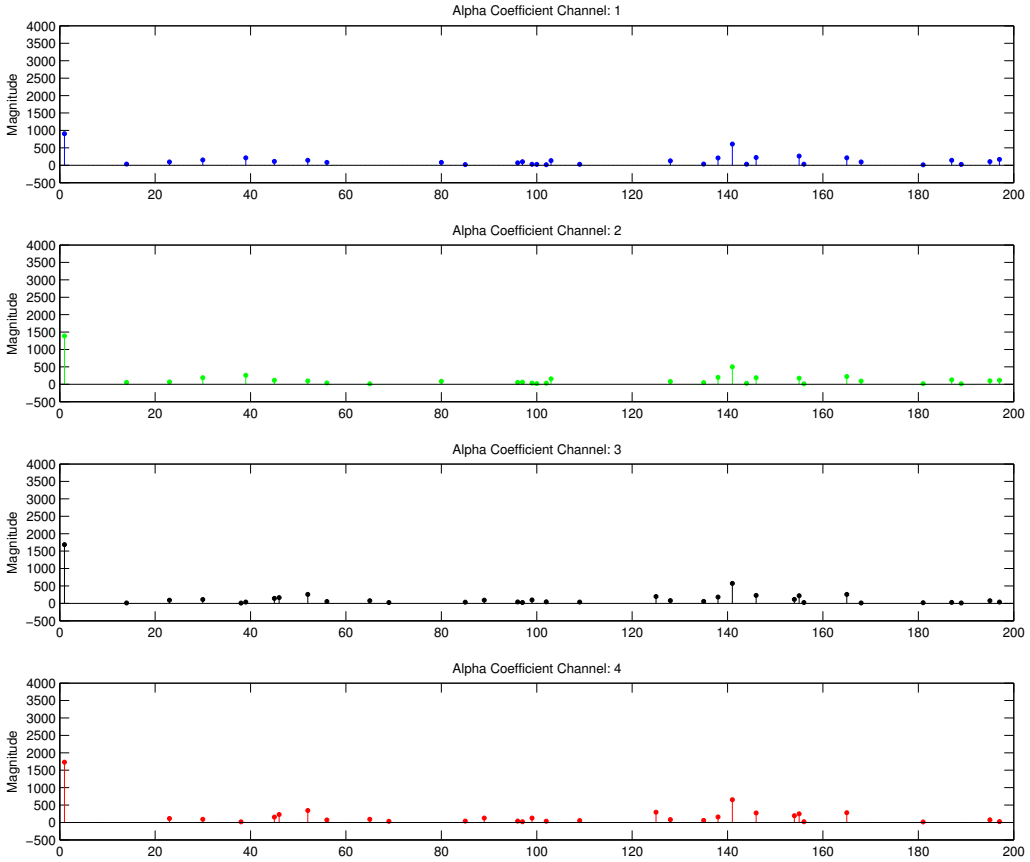
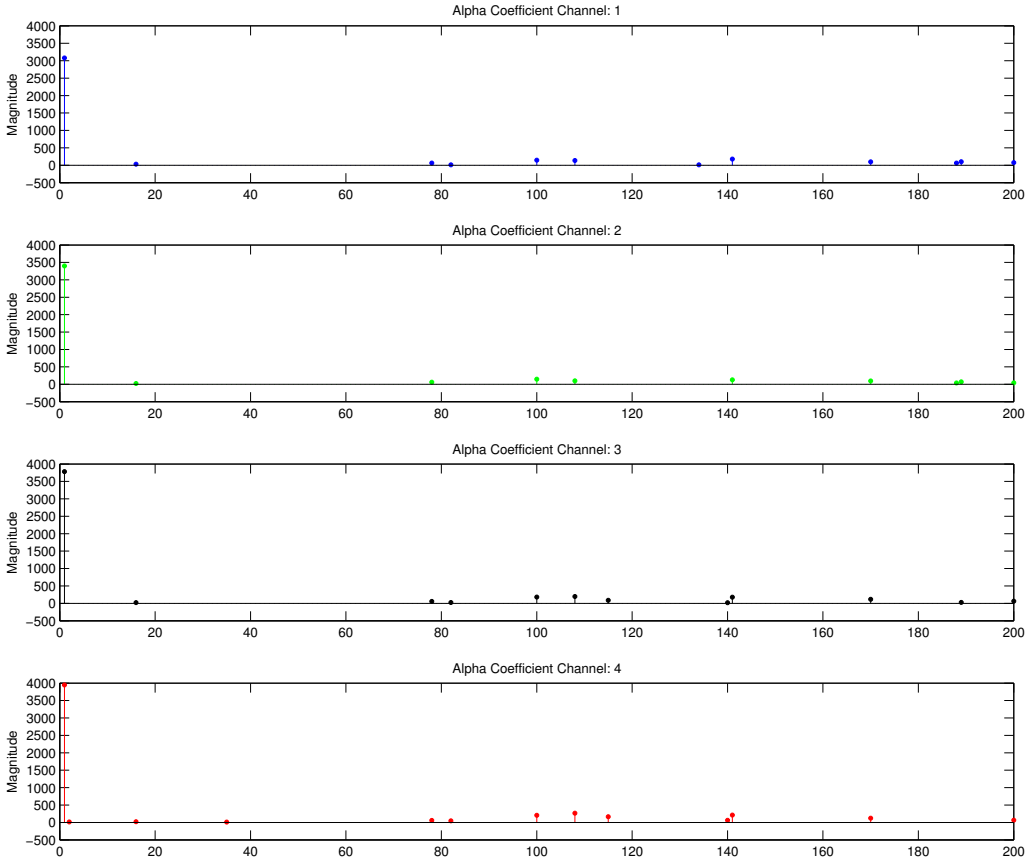


Figure B.18: NL8 Alphas



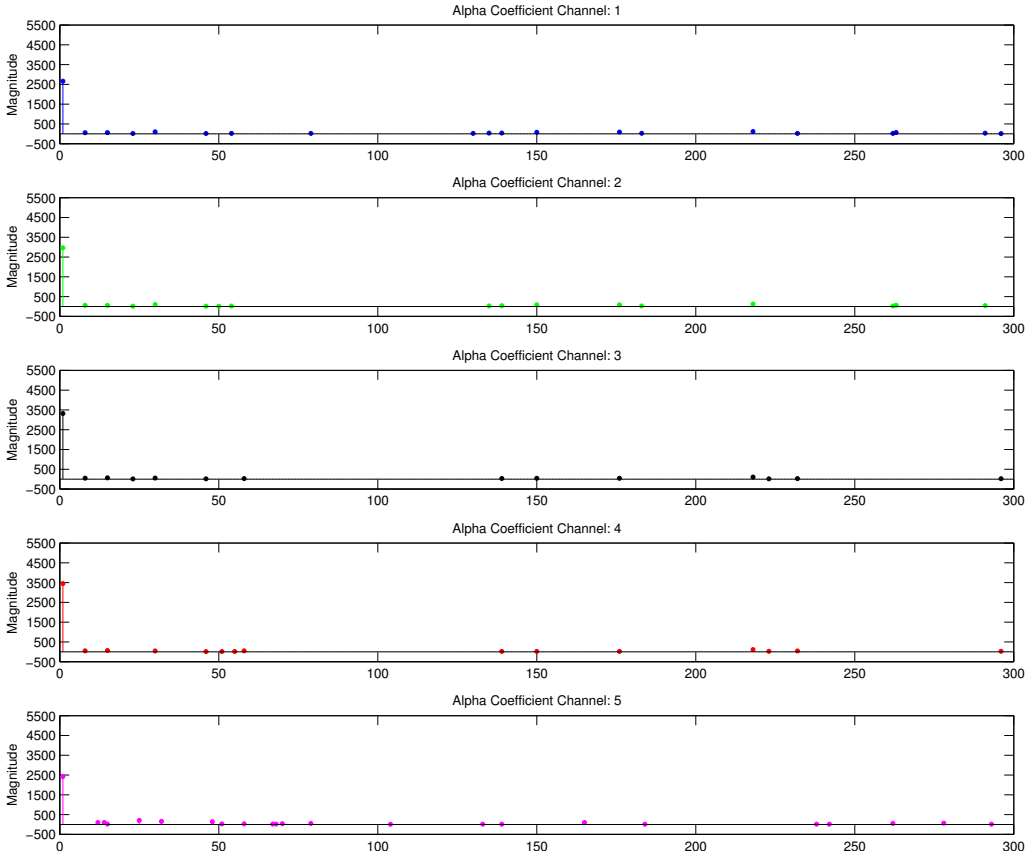
Appendix C

Supplementary Results: Dictionary Training

C.1 Alpha Coefficients

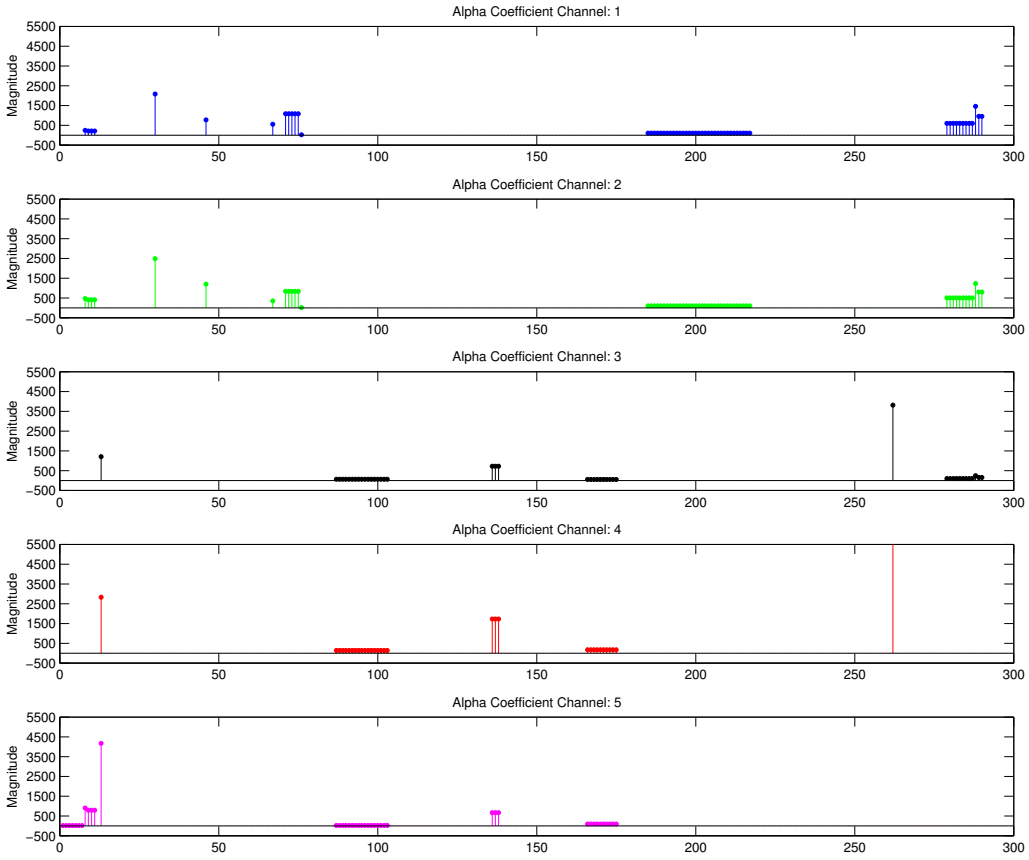
C.1.1 NNP Alpha Coefficients Prior to K-SVD

Figure C.1: NNP Alphas Prior to K-SVD



C.1.2 NNP Alpha Coefficients Post K-SVD

Figure C.2: NNP Alphas Post K-SVD



Bibliography

[1] Deepali A.Godse and Dattatraya S. Bormane. "Wavelet based image fusion using pixel based maximum selection rule". In: *International Journal of Engineering Science and Technology (IJEST)* 3.7 (2011).

[2] J Henkel J E Bare R Haydn G W Dalke. "Application of IHS color transform to the processing of multisensor data and image enhancement". In: *Proc. Int. Symp. Remote Sens. Arid and Semi-Arid Lands* 1 (1982), pp. 599–616.

[3] A. B. Kahle A. R. Gillespie and R. E. Walker. "Colour Enhancement of Highly Correlated Images". In: *Remote Sens. Environ.* 22.3 (1987), pp. 343–365.

[4] Shettigara, VK. "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set". In: *Photogrammetric Engineering and remote sensing* 58.5 (1992), pp. 561–567.

[5] B.V. Brower and C.A. Laben. *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*. US Patent 6,011,875. 2000.

[6] H.A. Aly and G. Sharma. "A Regularized Model-Based Optimization Framework for Pan-Sharpening". In: *Image Processing, IEEE Transactions on* 23.6 (2014), pp. 2596–2608. ISSN: 1057-7149.

[7] Xavier Otazu et al. "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods". In: *Geoscience and Remote Sensing, IEEE Transactions on* 43.10 (2005), pp. 2376–2385.

[8] Michael Moeller, Todd Wittman, and Andrea L Bertozzi. "Variational wavelet pansharpening". In: *CAM Report* (2008), pp. 08–81.

[9] Krista Amolins, Yun Zhang, and Peter Dare. "Wavelet based image fusion techniquesAn introduction, review and comparison". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 62.4 (2007), pp. 249–263.

[10] Shutao Li and Bin Yang. "A new pan-sharpening method using a compressed sensing technique". In: *Geosci. Remote Sensing, IEEE ...* 49.2 (2011), pp. 738–746.

[11] Xiao Xiang Zhu and Richard Bamler. "A sparse image fusion algorithm with application to pan-sharpening". In: *Geoscience and Remote Sensing, IEEE Transactions on* 51.5 (2013), pp. 2827–2836.

[12] XX Zhu, Sofya Spiridonova, and Richard Bamler. "A pan-sharpening algorithm based on joint sparsity". In: *Adv. Radar ...* 2 (2012).

[13] Grohnfeldt Claas Zhu Xiao Xiang and Richard Bamler. "Exploiting Joint Sparsity for Pan-Sharpening - The J-SparseFI Algorithm". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.6 (2014), pp. 797–829.

[14] M. Aharon, M. Elad, and A. Bruckstein. "K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *Signal Processing, IEEE Transactions on* 54.11 (2006), pp. 4311–4322. ISSN: 1053-587X.

[15] Michal Aharon. "Overcomplete Dictionaries for Sparse Representation of Signals". PhD thesis. Israel Institute of Technology, 2006.

[16] Jonathan Li Ming Cheng Cheng Wang. "Sparse Representation Based Pansharpening Using Trained Dictionary". In: *IEEE Geoscience and Remote Sensing Letters* 11.1 (2014), pp. 293–297.

[17] Claas Grohnfeldt, Xiao Xiang Zhu, and Richard Bamler. "The J-SparseFI-HM Hyperspectral Resolution Enhancement Method - Now Fully Automated". In: *Work. Hyperspectral Image Signal Process.* (2014), pp. 1–5.

[18] Sarah Cohen. "Compressive Sensing with Highly Coherent Dictionaries". MA thesis. One Shields Avenue, Davis, CA 95616: The University of California, Davis, 2012.

[19] Michael Elad, Mario AT Figueiredo, and Yi Ma. "On the role of sparse and redundant representations in image processing". In: *Proceedings of the IEEE* 98.6 (2010), pp. 972–982.

[20] Emmanuel J Candes, Justin K Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on pure and applied mathematics* 59.8 (2006), pp. 1207–1223.

[21] Julien Mairal, Michael Elad, and Guillermo Sapiro. "Sparse representation for color image restoration". In: *Image Processing, IEEE Transactions on* 17.1 (2008), pp. 53–69.

[22] Michael Elad and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries". In: *Image Processing, IEEE Transactions on* 15.12 (2006), pp. 3736–3745.

[23] John Wright et al. "Robust face recognition via sparse representation". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.2 (2009), pp. 210–227.

[24] Jianchao Yang et al. "Image super-resolution as sparse representation of raw image patches". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 2008, pp. 1–8.

[25] David L Donoho. "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution". In: *Communications on pure and applied mathematics* 59.6 (2006), pp. 797–829.

[26] Xiao Xiang Zhu, Claas Grohnfeldt, and Richard Bamler. "Collaborative sparse reconstruction for pan-sharpening". In: *Geosci. Remote Sens. Symp. (IGARSS), IEEE Int.* 1 (2013), pp. 0–3.

[27] Massimo Fornasier and Holger Rauhut. "Recovery algorithms for vector-valued data with joint sparsity constraints". In: *SIAM Journal on Numerical Analysis* 46.2 (2008), pp. 577–613.

[28] Massimo Fornasier and Holger Rauhut. "Iterative thresholding algorithms". In: *Applied and Computational Harmonic Analysis* 25.2 (2008), pp. 187–208.